# WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



### INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: Č12N 15/31, 1/21, C12P 21/02, C07K 14/33, A61K 38/16, 39/08

(I1) International Publication Number:

WO 98/07864

A1 (43) International Publication Date:

26 February 1998 (26.02.98)

(21) International Application Number:

PCT/GB97/02273

(22) International Filing Date:

22 August 1997 (22.08.97)

(30) Priority Data:

9617671.4 9625996.5 23 August 1996 (23.08.96)

GB 13 December 1996 (13.12.96) GB

Conrad, Padraig [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). FOSTER, Keith, Alan [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 OJG (GB).

(74) Agent: SCHLICH, George, William; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).

(60) Parent Application or Grant

(63) Related by Continuation

US Filed on

08/782,893 (CIP) 27 December 1996 (27.12.96)

(71) Applicants (for all designated States except US): MICRO-BIOLOGICAL RESEARCH AUTHORITY CAMR (CEN-TRE FOR APPLIED MICROBIOLOGY & RESEARCH) [GB/GB]; Porton Down, Salisbury, Wiltshire SP4 0JG (GB). THE SPEYWOOD LABORATORY LIMITED [GB/GB]; 14 Kensington Square, London W8 5HH (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SHONE, Clifford, Charles [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). QUINN,

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, I'R, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

#### Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RECOMBINANT TOXIN FRAGMENTS

#### (57) Abstract

A polypeptide has first and second domains which enable the polypeptide to be translocated into a target cell or which increase the solubility of the polypeptide, or both, and further enable the polypeptide to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis. The polypeptide thus combines useful properties of a clostridial toxin, such as a botulinum or tetanus toxin, without the toxicity associated with the natural molecule. The polypeptide can also contain a third domain that targets it to a specific cell, rendering the polypeptide useful in inhibition of exocytosis in target cells. Fusion proteins comprising the polypeptide, nucleic acids encoding the polypeptide and methods of making the polypeptide are also provided. Controlled activation of the polypeptide is possible and the polypeptide can be incorporated into vaccines and toxin assays.

### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

	•							
Al.	Albania	ES	Spain	LS	Lesotho	SI	Slovenia	
AM	Armenia	Fl	Finland	LT	Lithuania	SK	Slovakia	
AT	Austria	FR	France	1.0	Luxembourg	SN	Senegal	
AU	Australia	GA	Gabon	1.V	Latvia	<b>S7.</b>	Swaziland	
A7.	Azerbaijan	GB	United Kingdom	MC	Мопасо	TD	Chad	
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TC	Togo	
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan	
BF.	Belgium	GN	Guines	MK	The former Yugoslav	TM	Turkmenistan	
BF.	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey	
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago	
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine	
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda	
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America	
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan	
CF	Central African Republic	JР	Japan	NE	Niger	VN	Viet Nam	
CC	Congo	KE	Kenya	NL.	Netherlands	YU	Yugoslavia	
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe	
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand			
CM	Cameroon		Republic of Korea	PL	Poland			
CN	China	KR	Republic of Korea	PT	Portugal			
CU	Cuba	KZ.	Kazakstan	RO	Romania			
C7.	Czech Republic	LC	Saint Lucia	RU	Russian Federation			
DE	Germany	LI	Liechtenstein	SD	Sudan			
DK	Denmark	LK	Sri Lanka	· SE	Sweden			
EE	Estonia	LR	1.iberia	SG	Singapore			

#### RECOMBINANT TOXIN FRAGMENTS

This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the  $H_{\rm C}$  domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the  $H_{\rm N}$  domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The  $H_{\rm N}$  domain also has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the  $H_{\rm C}$ , amino acid residues 449-871 for the  $H_{\rm N}$  and residues 1-448 for the LC. Digestion with trypsin effectively degrades the  $H_{\rm C}$  domain of the BoNT/A to generate a non-toxic fragment designated  $LH_{\rm N}$ ,

which is no longer able to bind to and enter neurons (Fig. 1). The  $LH_N$  fragment so produced also has the property of enhanced solubility compared to both the parent holotoxin and the isolated LC.

It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

### (A) clostridial neurotoxin light chain:

-a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma - membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin.

#### (B) clostridial neurotoxin heavy chain H<sub>N</sub> domain:

- -a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.
- -the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.
- -the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.
- -the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.
- (C) clostridial neurotoxin heavy chain H<sub>c</sub> domain.
- -a portion of the heavy chain which is responsible for binding of the native

holotoxin to cell surface receptor(s) involved in the intoxicating action of clostridial toxin prior to internalisation of the toxin into the cell.

The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

However, due to its extreme toxicity, the handling of native toxin is hazardous.

The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

Accordingly, the invention provides a polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to neuronal exocytosis and wherein said second domain is adapted (i) to translocate the polypeptide into the cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into the cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of any clostridial neurotoxin precursor that can be converted into toxin by proteolytic action. Accordingly, the invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the H<sub>N</sub> of a clostridial toxin heavy chain, whilst lacking the functional aspects of a clostridial toxin H<sub>C</sub> domain.

For the purposes of the invention, the functional property or properties of the  $H_N$  of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter to a  $H_N$  domain or to the functions of a  $H_N$  domain are references to this property or properties. The second domain is not required to exhibit other properties of the  $H_N$  domain of a clostridial toxin heavy chain.

A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin-this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH<sub>423</sub>/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is-of-use if-solubility-is-imparted to a composition made up of that polypeptide and one or more other components by one or more of said other components.

The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has, indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from Aplysia (sea slug), Drosophila (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

The polypeptide of the invention may be obtained by expression of a recombinant nucleic acid, preferably a DNA, and is a single polypeptide, that is to say not

cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

In a polypeptide according to the invention, said first domain preferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasma-membrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in J. Biol. Chem., Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment (ii) a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between the LC and  $H_N$  components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

In an embodiment of the invention described in an example below, the toxin light

chain and the portion of the toxin heavy chain are of botulinum toxin typ. A. In a further embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain  $H_N$  portion or a fragment or variant of a clostridial toxin heavy chain  $H_N$  portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the  $H_N$  domain. Teachings of regions within the  $H_N$  responsible for its function are provided for example in Biochemistry 1995, 34, pages 15175-15181 and Eur. J. Biochem, 1989, 185, pages 197-203. The variant has a different sequence from the  $H_N$  domain or fragment, though it too retains the function of the  $H_N$  domain or fragment thereof. In embodiments of the invention, described below, it comprises (i) an N-terminal extension to a  $H_N$  domain or fragment, (iii) a C-terminal extension to a  $H_N$  domain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

These polypeptides of the invention are thus not composed of two or more polypeptides, linked for example by di-sulphide bridges into composite molecules. Instead, these polypeptides are single chains and are not active or their activity is

significantly reduced in an in vitro assay of neurotoxin endopeptidase activity.

Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated  $H_{\rm C}$  of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragm nt of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated  $H_{\rm C}$  of a clostridial toxin heavy chain.

In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated  $H_N$  of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and  $H_N$  sequences of botulinum toxin types A, B, C<sub>1</sub>, D, E, F and G.

The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic IgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then determines the target for a polypeptide - immunoglobulin complex. Alternatively, the

polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

As noted above, by proteolytic treatment, for example using trypsin; of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide. A third aspect of the invention provides a composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the clostridial toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*. The activity of the derivative preferably approaches that of natural toxin, and is thus preferably at least 30% and most preferably at least 60% of natural toxin. The overall endopeptidase activity of the composition will, of course, also be determined by the amount of the derivative that is present.

While it is known to treat naturally produced clostridial toxin to remove the H<sub>C</sub> domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

The invention enables production of the polypeptides and fusion proteins of the invention by recombinant means.

A fourth aspect of the invention provides à nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described above.

In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain. This recombinant product is designated  $LH_{423}/A$  (SEQ ID NO: 2).

In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polypeptide or a fusion protein and comprises nucleotides encoding residues 1-

1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated LH<sub>728</sub>/B (SEQ

ID NO: 20).

The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host.

The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, the fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

The  $LH_N/A$  derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal 1/2 of the heavy chain, the  $H_C$  domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recovered from *E. coli*, and from other recombinant expression hosts, is an inactive single chain peptide or if expression

hosts produce a processed, active polypeptide it is not a toxin. Endopeptidase activity of LH<sub>423</sub>/A, as assessed by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H<sub>C</sub> domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EP-A-0763131.

A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant  $LH_N$  derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus

of the LC at the putative trypsin sensitive region and also at the extreme C-t rminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

The  $LH_N$  enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin-serotyp or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a / to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

Following this nomenclature,

LH <sub>423</sub> /A	=	SEQ ID NO: 2, containing the entire L-chain and 423 amino acids of the H-chain of botulinum neurotoxin type A;
<sub>2</sub> LH <sub>423</sub> /A	=	a variant of this molecule, containing a two amino acid extension to the N-terminus of the L-chain;
<sub>2</sub> L <sub>/2</sub> H <sub>423</sub> /A		a further variant in which the molecule contains a two amino acid extension on the N-terminus of both the L-chain and the H-chain;
<sub>2</sub> L <sub>FXa/2</sub> H <sub>423</sub> /A	=	a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the H-chain component; and
<sub>2</sub> L <sub>FXa/2</sub> H <sub>423</sub> /A-IGF-1	=	a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gen for an embodiment of the invention designated LH<sub>423</sub>/A;

- Fig. 3 is a graph comparing activity of native toxin, trypsin generated "native"  $LH_N/A$  and an embodiment of the invention designated  ${}_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) in an *in vitro* peptide cleavage assay;
- Fig. 4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;
- Fig. 5 shows the transition region of an embodiment of the invention designated L/4H423/A illustrating insertion of four amino acids at the N-terminus of the H<sub>N</sub> sequence; amino acids coded for by the *Eco* 47 III restriction endonuclease cleavage site are marked and the H<sub>N</sub> sequence then begins ALN...;
- Fig. 6 shows the transition-region of an embodiment of the invention designated  $L_{FXa/3}H_{423}/A$  illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H-sequence; the N-terminal amino acid of the cleavage-activated  $H_N$  will be cysteine;
- Fig. 7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated  $L_{FXa/3}H_{423}/A$ -IGF-1, a fusion protein; the IGF-1 sequence begins at position  $G_{882}$ ;
- Fig. 8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated  $L_{FXa/3}H_{423}/A$ -CtxA14, a fusion protein; the C-terminal CtxA sequence begins at position  $Q_{882}$ ;
- Fig.9 shows the C-terminal portion of the amino acid sequence of an

embodiment of the invention designated  $L_{FXa/3}H_{423}/A-ZZ$ , a fusion protein; the C-terminal ZZ sequence begins at position  $A_{890}$  immediately after a genenase recognition site (underlined);

show schematic representations of manipulations of

Figs. 10 & 11 polypeptides of the invention; Fig. 10 shows LH<sub>423</sub>/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an Ig binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage sites R1, R2 and R3 and a C-terminal fusion peptide sequence;

Fig. 12 shows the trypsin sensitive activation region of a polypeptide of the invention;

shows Western blot analysis of recombinant LH<sub>107</sub>/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH<sub>107</sub>/B; panel B was probed with anti-T7 peptid tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lan 4 immunopurified LH<sub>107</sub>/B.

The sequence listing that accompanies this application contains the following sequences:-

SEQ ID NO:

Sequence

1

DNA coding for LH<sub>423</sub>/A

2	LH <sub>423</sub> /A
3	DNA coding for $_{23}LH_{423}/A$ ( $Q_2E,N_{26}K,A_{27}Y$ ), of which an N-terminal portion is shown in Fig. 4.
4	<sub>23</sub> LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y)
5	DNA coding for $_2LH_{423}/A$ ( $Q_2E,N_{26}K,A_{27}Y$ ), of which an N-
	terminal portion is shown in Fig.4
6	<sub>2</sub> LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y)
	1
7	DNA coding for native BoNT/A according to Binz et al
8	native BoNT/A according to Binz et al
9	DNA coding for L <sub>/4</sub> H <sub>423</sub> /A
10	L <sub>/4</sub> H <sub>423</sub> /A
11	DNA coding for L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A
12	L <sub>Fxa</sub> / <sub>3</sub> H <sub>423</sub> /A
13	DNA coding for L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-IGF-1
14	L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-IGF-1
15	DNA coding for L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-CtxA14
16	L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-CtxA14
17	DNA coding for L <sub>FXa/3</sub> H <sub>423</sub> /A-ZZ
18	L <sub>FXe/3</sub> H <sub>423</sub> /A-ZZ
19	DNA coding for LH <sub>728</sub> /B
20	LH <sub>728</sub> /B
21	DNA coding for LH <sub>417</sub> /B
22	LH <sub>417</sub> /B
23	DNA coding for LH <sub>107</sub> /B
24	LH <sub>107</sub> /B
25	DNA coding for LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y)
26	LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y)
27	DNA coding for LH <sub>417</sub> /B wherein the first 274 bases are

modified to have an E.coli codon bias

28

DNA coding for  $LH_{417}/B$  wherein bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide

#### Example 1

A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain (H<sub>c</sub>) of botulinum neurotoxin type A. This recombinant product is designated the LH<sub>423</sub>/A fragment (SEQ ID NO: 2).

### Construction of the recombinant product

The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an *E. coli* codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique *KpnI* restriction site. The remainder of the LH<sub>423</sub>/A coding sequence was PCR amplified from total chromosomal DNA from *Clostridium botulinum* and annealed to the synthetic portion of the gene.

The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polymerase amplified DNA (bases 914-1138 and 1976-2616) and the remainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The

assembled gen was then fully sequenced and cloned into a variety of *E.coli* plasmid vectors for expression analysis.

### Expression of the recombinant gene and recovery of protein product

The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZZ18, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

Currently, E. coli harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth Expression of the gene is then induced phase. isopropylthio-β-D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced temperature of 25°C. The cells are then harvested by centrifugation, are resuspended in a buffer solution containing 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH<sub>423</sub>/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art.

The recombinant GST-  $LH_{423}/A$  is purified by adsorption onto a commercially prepared affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant  $LH_{423}/A$  is recovered in the non-adsorbed material.

#### Construct variants

A variant of the molecule,  $LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of  $LH_{423}/A$  producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A.

Two further variants of the gene sequence that have been expressed and the corresponding products purified are  $_{23}LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and  $_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

In yet another variant a gene has been produced which contains a *Eco* 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in th gene representing the interface of the heavy and light chains in native neurotoxin, and provides the capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the *Eco* 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, L<sub>14</sub>H<sub>423</sub>/A (SEQ ID NO: 10), which contains an additional four amino acids between amino acids 448 and 449 of LH<sub>423</sub>/A at a position equivalent to the amino terminus of the

heavy chain of native BoNT/A.

A variant of the gene has been expressed, L<sub>FXa/3</sub>H<sub>423</sub>/A (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of L<sub>/4</sub>H<sub>423</sub>/A. The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated, and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention.

Variants of  $L_{FXa/3}H_{423}/A$  have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.'

Specific examples described are:

- (1)  $L_{FXa/3}H_{423}/A$ -IGF-1 (SEQ ID NO: 14), in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity:
- (2)  $L_{FXa/3}H_{423}/A$ -CtxA14 (SEQ ID NO: 16), in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and
- (3) L<sub>FXa/3</sub>H<sub>423</sub>/A-ZZ (SEQ ID NO: 18), in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cleavage site located between the end of the clostridial heavy chain sequence and the sequence coding for the binding

- 21 -

ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a genenase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing  $H_N$  function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.

It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.

Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.

The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.

Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single polypeptide

incorporating a domain equivalent to the light chain of botulinum neurotoxin type A and a domain equivalent to the  $H_N$  domain of the heavy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely GST, and a C-terminal extension providing a ligand binding domain, namely an IgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

### Assay of product activity

The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The  $_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of  $_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains.

This activity is dependent on proteolytic modification of the recombinant  $GST_{-2}LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600  $\mu$ g/ml) is incubated at 37°C for 10-50 minutes with trypsin (10  $\mu$ g/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na $_2$ HPO $_4$ , 1.8 mM KH $_2$ PO $_4$ , pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

<sub>2</sub>LH<sub>423</sub>/A is more stable in the presence of trypsin and more active in the *in vitro* 

peptide cleavage assay than is  $_{23}LH_{423}/A$ . Both variants, how ver, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the recombinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moieties as would be obvious to those skilled in the art.

#### Example 2

As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

The gene sequences relating to this example were all assembled and express dusing methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

A\_gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide,  $LH_{728}/B$  (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

A gene has also been assembled coding for a variant polypeptide,  $LH_{417}/B$  (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus

equivalent by amino acid homology to that at the carboxy-terminus of the h avy chain fragment in native  $LH_N/A$ .

A gene has also been assembled coding for a variant polypeptide,  $LH_{107}/B$  (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

#### **Construct Variants**

A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E.coli* codon usage bias.

For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B w re synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. In addition, terminal restriction endonucleas sites of the synthetic products were constructed to facilitate insertion of thes products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27). Similarly the sequence could be inserted into other genes of the examples.

Another variant sequence equivalent to nucleotides 691 to 1641 of SEQ ID NO: 21

, and employing non-native codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

## Example 3

An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-<sub>2</sub>LH<sub>423</sub>/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA11) reactive against a conformation dependent epitope on the native LH<sub>N</sub>/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na<sub>2</sub>HPO<sub>4</sub> 1.15 g/l, KH<sub>2</sub>PO<sub>4</sub> 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

For immunisation, 20 µg of GST-<sub>2</sub>LH<sub>423</sub>/A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to its derivative LH<sub>N</sub>/A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na<sub>2</sub>HPO<sub>4</sub> anhydrous 10 g/l, gelatin (Difco) 2 g/l, pH 6.5-6.6) containing a dilution range from 0.5 µg (5X10<sup>-6</sup> g) to 5 picograms (5X10<sup>-12</sup> g). Aliquots of 0.5 ml were injected into mice intraperitoneally and deaths recorded

over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti-  $GST_{-2}LH_{423}/A$  antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1. Neutralisation of botulinum neurotoxin in mice by guinea pig anti-GST-2LH<sub>423</sub>/A antiserum.

Botulinum Toxin/mouse														
Survivors On Day	0.5 <i>µ</i> g	0.005µg	0.0005 <b>µ</b> g	0.5ng	0.005ng	5pg	Contr I							
1	0	4	4	4	4	4	. 4							
2		4	4	4	4	. 4	4							
3	-	4	4	4	4	4	• 4							
4		4	4	4	4	4	. 4							

TABLE 2. Neutralisation of botulinum neurotoxin in mice by non-immune guinea pig antiserum.

			_					
Survivors On Day		0.5µg	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)
	1	. 0	0	0	0	0	2	4
	2			-	•	•	0	4
	3 .	•	-	-	•	•	•	. 4
	4		-		•	-		4

**Botulinum Toxin/mouse** 

### Example 4

Expression of recombinant LH<sub>107</sub>/B in E. coli.

As an exemplification of the expression of a nucleic acid coding for a  $LH_N$  of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide  $LH_{107}/B$  (SEQ ID

NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in  $E.\ coli$  BL21 (DE3) (New England BioLabs, Beverley, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for  $LH_{423}/A$ .

Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverley, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7-antibody-and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

#### SEQUENCE LISTING

### (1) GENERAL INFORMATION:

- (:) APPLICANT:
  - (A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY
  - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
  - (A) NAME: THE SPEYWOOD LABORATORY LIMITED
  - (B) STREET: 14 Kensington Square
  - (C) CITY: London
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): W8 5HH
  - (A) NAME: FOSTER; Keith Alan
  - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
  - (A) NAME: QUINN; Conrad Padraig
  - (B) STREET: Centre For Applied Microbiology And Research,
    Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
  - (A) NAME: SHONE; Clifford Charles
  - (B) STREET: Centre For Applied Microbiology And Research,
  - Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E). COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
- (ii) TITLE OF INVENTION: Recombinant Toxin Fragments
- (iii) NUMBER OF SEQUENCES: 28
- (iv) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
- (2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2616 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..2616

(xi)	SEQUENCE	DESCRIPTION:	SEQ	ID	NO:	1:	
------	----------	--------------	-----	----	-----	----	--

ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	Pro	GTA Val	AAC Asn 15	GGT		48
GTT Val	GAC Asp	ATT	GCC Ala 20	Tyr	ATC Ile	AAA Lys	ATT	CCA Pro 25	AAC Asn	'GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro		96
GTG Val	AAG Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT	CAT His	AAC Asn .40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg		144
GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu		192
GCA Ala 65	AAG Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	TCA Ser	TAC Tyr	TAC Tyr	GAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80		240
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC Tyr	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC Thr	AÀA Lys	TTA Leu	TTC Phe 95	GAG Glu		288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val		336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT - Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys		384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr		432 .
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160		480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr		528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	•	576 ·
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu		624
					Ala	ACT ( Thr / 215				Val							672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	Gly	CAT His 230	CGT ( Arg :	CTG Leu	TAT Tyr	Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240		720

CGC	GTG Val	TTC Phe	AAG Lys	GTI Val 245	Asn	ACC Thr	AAC . Asn	GCC	TAC Tyr 250	Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	<b>768</b>
GA/ Glu	GTA Val	AGC Ser	TTC Phe 260	Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TT1 Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAC Lys	TTT Phe 290	Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	Lys 320	960
TAT	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392
AGT Ser 465	CCT Pro	TCA Ser	GAA Glu	Asp	AAT Asn 470	TTT Phe	ACT Thr	AAT Asn	GAT Asp	CTA Leu 475	AAT Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu 480	1440
ATT Ile	ACA Thr	TCT Ser	GAT . Asp	ACT Thr 485	AAT Asn	ATA Ile	GAA Glu	GCA Ala	GCA Ala 490	GAA Glu	GAA Glu	AAT Asn	ATT Ile	AGT Ser 495	TTA Leu	1488
GAT Asp	TTA Leu	Ile	CAA Gln 500	CAA Gln	TA <b>T</b> Tyr	TAT Tyr	Leu	ACC Thr 505	TTT Phe	AAT Asn	TTT Phe	GAT Asp	AAT Asn 510	GAA Glu	CCT Pro	1536

GAA Glu	AAT Asn	ATT Ile 515	TCA Ser	ATA Ile	GAA Glu	AAT Asn	CTT Leu 520	TCA Ser	AGT Ser	GAC Asp	ATT Ile	ATA Ile 525	GGC Gly	CAA Gln	TTA Leu		1584
GAA Glu	CTT Leu 530	ATG Met	CCT Pro	AAT Asn	ATA Ile	GAA Glu 535	AGA Arg	TTT Phe	CCT Pro	AAT Asn	GGA Gly 540	AAA Lys	AAG Lys	TAT	GAG Glu		1632
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT His	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560		1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu		1728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys		1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	Trp	GTA Val	GAA Glu		1824
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr		1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	•	1920
TTA Leu	ĀAT Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	GAT Asp 650	GAT Asp	TTT Phe	GTA Val	GGT Gly	GCT Ala 655	TTA Leu		1968
ATA Ile	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT Ile	GCA Ala		2016
ATA Ile	CCT Pro	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys		2064
GTT Val	CTA Leu 690	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	TAA naA	GAA Glu		2112
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720		2160
GTT Val	AAT Asn	ACA Thr	CAG Gln	ATT Ile 725	GAT Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu		2208
GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	TAA Asn		2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp		2304
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA Ile		2352

AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2400
			GGT Gly													2448
			TTA Leu 820													2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2544
			CAG Gln													2592
			GAA Glu				TAA				ı					2616

#### (2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 872 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 185 180 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 230 235 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 345 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 395 390 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 455 Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 505 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 520

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 555 560

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu
565 570 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 595

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 . 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655

Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685

Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700

Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 715 720

Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735

Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750

Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
755 760 765

Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 780

Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785 790 795 800

Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 860

Thr Phe Thr Glu Tyr Ile Lys \*865 870

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2685 base pairs (B) TYPE: nucleic acid

(C) STRANDEDNESS: double (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..2685

(xi)	SEQUENCE	DESCRIPTION:	SEQ	ID	NO:	3:
------	----------	--------------	-----	----	-----	----

GGA Gly 1	TCC Ser	CCA Pro	GGA Gly	ATT Ile 5	CAT His	ATG Met	ACG Thr	TCG Ser	ACG Thr 10	CGT Arg	CTG Leu	CAG Gln	AAG Lys	CTT Leu 15	CTA Leu		48
GAA Gļu	TTC Phe	GAG Glu	CTC Leu 20	CCG Pro	GGT Gly	ACC Thr	ATG Met	GAG Glu 25	TTC Phe	GTG Val	AAC Asn	AAG Lys	CAG Gln 30	TTC Phe	AAC Asn		96
TAT Tyr	AAG Lys	GAC Asp 35	CCT Pro	GTA Val	AAC Asn	GGT Gly	GTT Val 40	GAC Asp	ATT Ile	GCC Ala	TAC Tyr	ATC Ile 45	AAA Lys	ATT Ile	CCA Pro		144
AAG Lys	TAC Tyr 50	GGC Gly	CAG Gln	ATG Met	CAG Gln	CCG Pro 55	GTG Val	AAG Lys	GCT Ala	TTC Phe	AAG Lys 60	ATT Ile	CAT His	AAC Aşn	AAA Lys		192
ATC Ile 65	TGG Trp	GTT Val	ATT Ile	CCG Pro	GAA Glu 70	CGC Arg	GAT Asp	ACA Thr	TTT Phe	ACG Thr 75	AAC Asn	CCG Pro	GAA Glu	GAA Glu	GGA Gly 80	•	240
GAC Asp	TTG Leu	AAC Asn	CCG Pro	CCG Pro 85	CCG Pro	GAA Glu	GCA Ala	AAG Lys	CAG Gln 90	GTG Val	CCA Pro	GTT Val	TCA Ser	TAC Tyr 95	TAC Tyr		288
GAT Asp	TCA Ser	ACC Thr	TAT Tyr 100	CTG Leu	AGC Ser	ACA Thr	GAC Asp	AAC Asn 105	GAG Glu	AAG Lys	GAT Asp	AAC Asn	TAC Tyr 110	CTG Leu	AAG Lys		336
GGA Gly	GTG Val	ACC Thr 115	AAA Lys	TTA Leu	TTC Phe	GAG Glu	CGT Arg 120	ATT Ile	TAT Tyr	TCC Ser	ACT Thr	GAC Asp 125	CTG Leu	GGC Gly	CGT Arg		384
ATG Met	CTG Leu 130	CTG Leu	ACC Thr	TCA Ser	ATC Ile	GTC Val 135	CGC Arg	GGA Gly	ATC Ile	CCA Pro	TTT Phe 140	TGG Trp	GGT Gly	GGC Gly	AGT Ser		432
ACC Thr 145	ATT Ile	GAC Asp	ACG Thr	GAG Glu	TTG Leu 150	AAG Lys	GTT Val	ATT Ile	GAC Asp	ACT Thr 155	AAC Asn	TGC Cys	ATT Ile	AAC Asn	GTG Val 160	•	480
ATC Ile	CAA Gln	CCA Pro	GAC Asp	GGT Gly 165	AGC Ser	TAC Tyr	AGA Arg	TCT Ser	GAA Glu 170	GAA Glu	CTT Leu	AAC Asn	CTC Leu	GTA Val 175	ATC Ile		528
ATC Ile	GGG Gly	CCC Pro	TCC Ser 180	GCG Ala	GAC Asp	ATT Ile	ATC Ile	CAG Gln 185	TTT Phe	GAG Glu	TGC Cys	AAG Lys	AGC Ser 190	TTT Phe	GGC		576
CAC His	GAA Glu	GTG Val 195	TTG Leu	AAC Asn	CTG Leu	ACG Thr	CGT Arg 200	AAC Asn	GGT Gly	TAC Tyr	GGC Gly	TCT Ser 205	ACT Thr	CAG Gln	TAC Tyr		624

ATT Ile	CGT Arg 210	TTC Phe	AGC Ser	CCA Pro	GAC Asp	TTC Phe 215	ACG Thr	TTC Phe	GGT Gly	TTC Phe	GAG Glu 220	GAG Glu	AGC Ser	CTG Leu	GAG Glu	672
GTT Val 225	GAT Asp	ACC Thr	AAC Asn	CCG Pro	CTG Leu 230	TTG Leu	GGT Gly	GCA Ala	GGC Gly	AAG Lys 235	TTC Phe	GCA Ala	ACT Thr	GAT Asp	CCA Pro 240	720
GCG Ala	GTG Val	ACC Thr	CTG Leu	Ala	CAC	Glu	CTG Leu	ATC Ile	CAC His 250	GCC Ala	GGT Gly	CAT His	CGT Aŗg	CTG Leu 255	TAT Tyr	768
GGC Gly	ATT Ile	GCG Ala	ATT Ile 260	AAC Asn	CCG Pro	AAC Asn	CGC	GTG Val 265	TTC Phe	AAG Lys	GTT Val	AAC Asn	ACC Thr 270	AAC Asn	GCC Ala	816
TAC Tyr	TAC Tyr	GAG Glu 275	ATG Met	AGT Ser	GGT Gly	TTA Leu	GAA Glu 280	GTA Val	AGC Ser	TTC Phe	GAG Glu	GAA Glu 285	CTG Leu	CGC Arg	ACG Thr	864
TTC Phe	GGT Gly 290	GGC Ģly	CAT His	GAT Asp	GCG Ala	AAG Lys 295	TTT Phe	ATC Ile	GAC Asp	AGC Ser	TŢG Leu 300	CAG Gln	GAG Glu	AAC Asn	GAG Glu	912
TTC Phe 305	CGT Arg	CTG Leu	TAC Tyr	TAC Tyr	TAC Tyr 310	AAC Asn	AAG Lys	TTT Phe	AAA Lys	GAT Asp 315	ATT Ile	GCA Ala	AGT Ser	ACA Thr	CTG Leu 320	960
AAC Asn	AAG Lys	GCT Ala	AAG Lys	TCC Ser 325	ATT Ile	ĠTG Val	GGT Gly	ACC Thr	ACT Thr 330	GCT Ala	TCA Ser	TTA Leu	CAG Gln	TAT Tyr 335	ATG Met	1008
AAA Lys	AAT Asn	GTT Val	TTT Phe 340	AAA Lys	GAG Glu	AAA Lys	TAT Tyr	CTC Leu 345	CTA Leu	TCT Ser	GAA Glu	GAT Asp	ACA Thr 350	TCT Ser	GGA Gly	1056
AAA Lys	TTT Phe	TCG Ser 355	GTA Val	GAT Asp	AAA Lys	TTA Leu	AAA Lys 360	TTT Phe	GAT Asp	AAG Lys	TTA Leu	TAC Tyr 365	AAA Lys	ATG Met	TTA Leu	1104
ACA Thr	GAG Glu 370	ATT Ile	TAC Tyr	ACA Thr	GAG Glu	GAT Asp 375	AAT Asn	TTT Phe	GTT Val	AAG Lys	TTT Phe 380	TTT Phe	AAA Lys	GTA Val	CTT Leu	1152
AAC Asn 385	AGA Arg	aaa Lys	ACA Thr	TAT Tyr	TTG Leu 390	TAA Asn	TTT Phe	GAT Asp	AAA Lys	GCC Ala 395	GTA Val	TTT Phe	AAG Lys	ATA Ile	AAT Asn 400,	1200
ATA Ile	GTA Val	CCT Pro	AAG Lys	GTA Val 405	AAT Asn	TAC Tyr	ACA Thr	ATA Ile	TAT Tyr 410	GAT Asp	GGA Gly	TTT Phe	AAT Asn	TTA Leu 415	AGA Arg	1248
AAT Asn	ACA Thr	AAT Asn	TTA Leu 420	GCA Ala	GCA Ala	AAC Asn	TTT Phe	AAT Asn 425	GGT Gly	CAA Gln	TAA Asn	ACA Thr	GAA Glu 430	ATT Ile	AAT Asn	1296
AAT Asn	ATG Met	AAT Asn 435	TTT Phe	ACT Thr	AAA Lys	CTA Leu	AAA Lys 440	AAT Asn	TTT Phe	ACT Thr	GGA Gly	TTG Leu 445	TTT Phe	GAA Glu	TTT Phe	1344
TAT Tyr	AAG Lys 450	TTG Leu	CTA Leu	TGT Cys	GTA Val	AGA Arg 455	GGG Gly	ATA Ile	ATA Ile	ACT Thr	TCT Ser 460	AAA Lys	ACT Thr	AAA Lys	TCA Ser	1392
TTA Leu 465	GAT Asp	AAA Lys	GGA Gly	TAC Tyr	AAT Asn 470	AAG Lys	GCA Ala	TTA Leu	AAT Asn	GAT Asp 475	TTA Leu	TGT Cys	ATC Ile	AAA Lys	GTT Val 480	1440

AA As	T AA' n As:	T TG n Tr	G GA p As	C TT p Le 48	u Pho	T TT:	r AG′ ≘ Sei	r CC	TC/ Ser 490	r Gli	A GA:	T AAT P Asi	TT:	T AC:	T AAT C Asn	1488	
GA As	T CT	A AA u As	T AA n Ly: 500	s GI	A GAI y Glu	A GAZ J Glu	ATT	Third Social Property of the Control	: Sei	GAT Asp	AC:	T AAT	ATA Ile 510	Glu	A GCA 1 Ala	1536	
Ala	a Gli	51!	u Ası 5	1 116	e Ser	Leu	Asp 520	Leu )	lle	Gln	Glr	Tyr 525	Туг	Lei	ACC Thr	1584	
Phe	530	n Phe	e Asp	) Asr	) GIV	535	Glu	Asn	Ile	Ser	11e 540	Glu	Asn	Leu		1632	
Ser 545	Asp	) Ile	r ATA e Ile	e Gly	550	Leu	Glu	Leu	Met	Pro 555	Asn	lle	Glu	Arg	Phe 560	1680	
Pro	Asn	Gly	AAA Lys	565	Tyr	Glu	Leu	Asp	Lys 570	Tyr	Thr	Met	Phe	His 575	Tyr	1728	
Leu	Arg	Ala	CAA Gln 580	Glu	Phe	Glu	His	Gly 585	ŗ'ns	Ser	Arg	Ile	Ala 590	Leu	Thr	1776	•
Asn	Ser	Val 595		Glu	Ala	Leu	Leu 600	Asn	Pro	Ser	Arg	Val 605	Tyr	Thr	Phe	1824	
Phe	Ser 610	Ser	GAC Asp	Tyr	Val	Lys 615	Lys	Val	Asn	Lys	Ala 620	Thr	Glu	Ala	Ala	1872	
Met 625	Phe	Leu	GGC Gly	Trp	Val 630	Glu	Gln	Leu	Val	Tyr 635	Asp	Phe	Thr	Asp	Glu 640	1920	
Thr	Ser	Glu	GTA Val	Ser 645	Thr	Thr	Asp	Lys	Ile 650	Ala	Asp	Ile	Thr	11e 655	Ile	1968	
Ile	Pro	Tyr	ATA Ile 660	Gly	Pro	Ala	Leu	Asn 665	Ile	Gly	Asn	Met	Leu 670	Tyr	Lys	2016	
GAT Asp	GAT Asp	TTT Phe 675	GTA Val	GGT Gly	GCT Ala	TTA Leu	ATA Ile 680	TTT Phe	TCA Ser	GGA Gly	GCT Ala	GTT Val 685	ATT Ile	CTG Leu	TTA Leu	2064	
GAA Glu	TTT Phe 690	ATA Ile	CCA Pro	GAG Glu	Ile	GCA Ala 695	ATA Ile	CCT Pro	GTA Val	Leu	GGT Gly 700	ACT Thr	TTT Phe	GCA Ala	CTT Leu	2112	
GTA Val 705	TCA Ser	TAT Tyr	ATT Ile	Ala	AAT Asn 710	AAG Lys	GTT Val	CTA Leu	ACC Thr	GTT Val 715	CAA Gln	ACA Thr	ATA Ile	GAT Asp	AAT Asn 720	2160	
GCT Ala	TTA Leu	AGT Ser	AAA Lys	AGA Arg 725	TAA neA	GAA . Glu	AAA Lys	Trp	GAT Asp 730	GAG Glu	GTC Val	TAT Tyr	AAA Lys	TAT Tyr 735	ATA Ile	2208	
GTA Val	ACA Thr	AAT Asn	TGG Trp 740	TTA   Leu .	GCA . Ala :	AAG ( Lys	Val .	AAT Asn 745	ACA Thr	CAG . Gln	ATT Ile	Asp	CTA Leu 750	ATA Ile	AGA Arg	2256	

	Lys						AAG Lys		2304
							AAT Asn		2352
		Ile	Asp				TCT Ser	,	2400
							TCT Ser 815		2448
							TTA Leu		2496
							TAT Tyr		2544
							AAA Lys		2592
 							TAC Tyr		2640
	AGA Arg						TAA * 895		2685

### (2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 895 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu 1 5 10 15

Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn 20  $^{-1}$ 

Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro 35 40

Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys 50 55 60

Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly 65 70 75 80

Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr 85 90 95

Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys
100 105 110

Gly	Val	Thr 115	Lys	Leu	ı Phe	Glu	Arg 120	Ile	Tyr	Ser	Thr	Asp 125		Gly	Arg
Met	Leu 130		Thr	Ser	Ile	Val 135	Arg	Gly	Ile	Pro	Phe 140		Gly	Gly	Ser
Thr 145		Asp	Thr	Glu	Leu 150		Val	Ile	Asp	Thr 155		Cys	Ile	Asn	Val 160
Ile	Gln	Pro	Asp	Gly 165		Tyr	Arg	Ser	Glu 170		Leu	Asn	Leu	Val 175	Ile
Ile	Gly	Pro	Ser 180		Asp	Ile	Ile	Gln 185		Glu	Cys	Lys	Ser 190	Phe	Gly
Hıs	Glu	Val 195	Leu	Asn	Leu	Thr	Arg 200	Asn	Gly	Tyr	Gly	Ser 205	Thr	Gln	Tyr
Ile	Arg 210	Phe	Ser	Pro	Asp	Phe 215	Thr	Phe	Gly	Phe	Glu 220	Glu	Ser	Leu	Glu
Val 225	Asp	Thr	Asn	Pro	Leu 230	Leu	Gly	Ala	Gly	Lys 235	Phe	Ala	Thr	Asp	Pro 240
Ala	Val	Thr	Leu	Ala 245	His	Glu	Leu	Ile	His 250	Ala	Gly	His	Arg	Leu 255	Tyr
Gly	Ile	Ala	11e 260	Asn	Pro	Asņ	Arg	Val 265	Phe	Lys	Val	Asn	Thr 270	Asn	Ala
Tyr 	Tyr	Glu 275	Met 	Ser	Gly	Leu	Glu 280	Val	Ser	Phe	Glu	Glu 285	Leu	Arg,	Thr
Phe	Gly 290	Gly	His	Asp	Ala	Lys 295	Phe	Ile	Asp	Ser	Leu 300	Gln	Glu	Asn	Glu
Phe 305	Arg	Leu	Tyr	Tyr	Tyr 310	Asn	Lys	Phe	Lys	Asp 315	Ile	Ala	Ser	Thr	Leu 320
Asn	Lys	Ala	Lys	Ser 325	Ile	Val	Gly	Thr	Thr 330	Ala	Ser	Leu	Gln	Tyr 335	Met
Lys	Asn	Val	Phe 340	Lys	Glu	Lys	Tyr	Leu 345	Leu	Ser	Glu	Asp	Thr 350	Ser	Gly
		355	Val				360					365	_		
Thr	Glu 370	Ile	Tyr	Thr	Glu	Asp 375	Asn	Phe	Val	Lys	Phe 380	Phe	Lys	Val	Leu
Asn 385	Arg	Lys	Thr	Tyr	Leu 390	Asn	Phe	Asp	Lys	Ala 395	Val	Phe	Lys	Ile	Asn 400
			Lys	405					410					415	
			Leu 420					425					430		
		435	Phe				440					445			
Tyr	Lys 450	Leu	Leu	Cys		Arg 455	Gly	lle	Ile	Thr	Ser 460	Lys	Thr	Lys	Ser

Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val 470 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 490 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr . . . . 520 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 535 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 550 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 570 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 600 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 695 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 725 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 745 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 760 Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 795 Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 820 825 Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 855 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 885 890 895

# (2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2622 base pairs
  - (B) TYPE: nucleic acid (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:

  - (A) NAME/KEY: CDS
    (B) LOCATION:1..2622

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

						AAC Asn										48	,
						TAC Tyr										96	,
						AAG Lys										144	
						AAC Asn 55										192	
						CCA Pro										240	
						GAT Asp										288	,
TTC Phe	GAG Glu	CGT Arg	ATT Ile 100	TAT Tyr	TCC Ser	ACT Thr	GAC Asp	CTG Leu 105	GGC Gly	CGT Arg	ATG Met	CTG Leu	CTG Leu 110	ACC Thr	TCA Ser	336	į
ATC Ile	GTC Val	CGC Arg 115	GGA Gly	ATC Ile	CCA Pro	TTT Phe	TGG Trp 120	GGT Gly	GGC Gly	AGT Ser	ACC Thr	ATT Ile 125	GAC Asp	ACG Thr	GAG Glu	384	•
TTG Leu	AAG Lys 130	GTT Val	ATT Ile	GAC Asp	ACT Thr	AAC Asn 135	TGC Cys	ATT Ile	AAC Asn	GTG Val	ATC Ile 140	CAA Gln	CCA Pro	GAC Asp	GGT Gly	432	:

	TAC					Leu					Ile					480
	ATT Ile				Glu										Asn	528
	ACG Thr			Gly												<b>576</b>
GAC Asp	TTC Phe	ACG Thr 195	TTC Phe	GGT Gly	TTC Phe	GAG Glu	GAG Glu 200	AGC Ser	CTG Leu	GAG Glu	GTT Val	GAT Asp 205	ACC Thr	AAC Asn	CCG Pro	624
CTG Leu	TTG Leu 210	GGT Gly	GCA Ala	GGC Gly	Lys	TTC Phe 215	GCA Ala	ACT	GAT Asp	CCA Pro	GCG Ala 220	GTG Val	ACC Thr	CTG Leu	GCA Ala	672
CAC His 225	GAG Glu	CTG Leu	ATC Ile	CAC His	GCC Ala 230	GGT Gly	CAT His	CGT	CTG Leu	TAT Tyr 235	GGC Gly	ATT Ile	GCG Ala	ATT Ile	AAC Asn 240	720
CCG Pro	AAC Asn	CGC Arg	GTG Val	TTC Phe 245	AAG Lys	GTT Val	AAC Asn	ACC Thr	AAC Asn 250	GCC Ala	TAC Tyr	TAC Tyr	GAG Glu	ATG Met 255	AGT Ser	768
GGT Gly	TTA Leu	GAA Glu	GTA Val 260	AGC Ser	TTC Phe	GAG Glu	GAA Glu	CTG Leu 265	CGC Arg	ACG Thr	TTC Phe	GGT Gly	GGC Gly 270	CAT His	GAT Asp	816
GCG Ala	AAG Lys	TTT Phe 275	ATC Ile	GAC Asp	AGC Ser	TTG Leu	CAG Gln 280	GAG Glu	AAC Asn	GAG Glu	TTC Phe	CGT Arg 285	CTG Leu	TAC Tyr	TAC Tyr	864
TAC Tyr	AAC Asn 290	AAG Lys	TTT Phe	AAA Lys	GAT Asp	ATT Ile 295	GCA Ala	AGT Ser	ACA Thr	CTG Leu	AAC Asn 300	AAG Lys	GCT Ala	AAG Lys	TCC Ser	912
ATT Ile 305	GTG Val	GGT Gly	ACC Thr	ACT Thr	GCT Ala 310	TCA Ser	TTA Leu	CAG Gln	TAT Tyr	ATG Met 315	AAA Lys	AAT Asn	GTT Val	TTT Phe	AAA Lys 320	960
GAG Glu	AAA Lys	TAT Tyr	CTC Leu	CTA Leu 325	TCT Ser	GAA Glu	GAT Asp	ACA Thr	TCT Ser 330	GGA Gly	AAA Lys	TTT Phe	TCG Ser	GTA Val 335	GAT Asp	1008
AAA Lys	TTA Leu	AAA Lys	TTT Phe 340	GAT Asp	AAG Lys	TTA Leu	TAC Tyr	AAA Lys 345	ATG Met	TTA Leu	ACA Thr	GAG Glu	ATT Ile 350	TAC Tyr	ACA Thr	1056
GAG Glu	GAT Asp	AAT Asn 355	TTT Phe	GTT Val	AAG Lys	TTT Phe	TTT Phe 360	AAA Lys	GTA Val	CTT Leu	AAC Asn	AGA Arg 365	AAA Lys	ACA Thr	TAT Tyr	1104
TTG Leu	AAT Asn 370	TTT Phe	GAT Asp	AAA Lys	Ala	GTA Val 375	TTT Phe	AAG Lys	ATA Ile	TAA Asn	ATA Ile 380	GTA Val	CCT Pro	AAG Lys	GTA Val	1152
AAT Asn 385	TAC Tyr	ACA Thr	ATA Ile	Tyr	GAT Asp 390	GGA Gly	TTT Phe	AAT Asn	TTA Leu	AGA Arg 395	AAT Asn	ACA Thr	AAT Asn	TTA Leu	GCA Ala 400	1200
GCA Ala	AAC Asn	TTT Phe	Asn	GGT Gly 405	CAA Gln	AAT Asn	ACA Thr	GAA Glu	ATT Ile 410	AAT Asn	AAT Asn	ATG Met	AAT Asn	TTT Phe 415	ACT Thr	1248

AAA Lys	CTA Leu	AAA Lys	AAT Asn 420	Phe	ACT	GGA Gly	TTG Leu	TTT Phe 425	Glu	TTT Phe	TAT Tyr	AAG Lys	TTG Leu 430	CTA Leu	TGT Cys		1296
GTA Val	AGA Arg	GGG Gly 435	ATA Ile	ATA Ile	ACT	TCT	AAA Lys 440	Thr	AAA Lys	TCA Ser	TTA Leu	GAT Asp 445	AAA Lys	GGA Gly	TAC Tyr		1344
		Ala	TTA Leu		Asp												1392
			CCT Pro											Lys			1440
			ACA Thr		Asp												1488
AGT Ser	TTA Leu	GAT Asp	TTA Leu 500	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 505	TTA Leu	ACC Thr	TTT, Phe	AAT Asn	TTT Phe 510	GAT Asp	AAT Asn	,	1536
			AAT Asn														1584
			CTT														1632
			GAT Asp														1680
TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 565	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 570	ACA Thr	AAT Asn	TCT Ser	GTT Val	AAC Asn 575	GAA Glu		1728
			AAT Asn 580														1776
GTA Val	AAG Lys	AAA Lys 595	GTT Val	AAT Asn	AAA Lys	GCT Ala	ACG Thr 600	GAG Glu	GCA Ala	GCT Ala	ATG Met	TTT Phe 605	TTA Leu	GGC Gly	TGG Trp		1824
GTA Val	GAA Glu 610	CAA Gln	TTA Leu	GTA Val	TAT Tyr	GAT Asp 615	TTT Phe	ACC Thr	GAT Asp	GAA Glu	ACT Thr 620	AGC Ser	GAA Glu	GTA Val	AGT Ser		1872
ACT Thr 625	ACG Thr	GAT Asp	AAA Lys	ATT Ile	GCG Ala 630	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 635	ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 640		1920
CCT Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 645	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 650	AAA Lys	GAT Asp	GAT Asp	TTT Phe	GTA Val 655	GGT Gly		1968.
GCT Ala	TTA Leu	ATA Ile	TTT Phe 660	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 665	CTG Leu	TTA Leu	GAA Glu	TTT Phe	ATA Ile 670	CCA Pro	GAG Glu		2016
ATT Ile	GCA Ala	ATA Ile 675	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 680	TTT Phe	GCA Ala	CTT Leu	GTA Val	TCA Ser 685	TAT Tyr	ATT Ile	GCG Ala		2064

AA: . Asi	r AAC n Lys 690	Va.	r CTA	ACC Thr	GTI Val	CAA Gln 695	Thr	ATA	GAI Asp	AAT Asn	GCT Ala 700	Leu	AGT Ser	AAA Lys	AGA Arg		2112
AAT Asi 705	ı Glu	AAJ Lys	TGG Trp	GAT Asp	GAG Glu 710	Val	TAT	AAA Lys	TAT Tyr	ATA Ile 715	GTA Val	ACA Thr	AAT Asn	TGG	TTA Leu 720		2160
GCA Ala	AAG Lys	GT1 Val	AAT Asn	ACA Thr 725	Gln	ATT ,Ile	GAT Asp	CTA Leu	ATA Ile 730	AGA Arg	AAA Lys	AAA Lys	ATG Met	AAA Lys' 735	GAA Glu		2208
GCT Ala	TTA Leu	GAA Glu	AAT Asn 740	Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 745	AAG Lys	GCT Ala	ATA Ile	ATA Ile	AAC Asn 750	TAT Tyr	CAG Gln		2256
			TAT													ı	2304
GAT Asp	GAT Asp 770	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 775	AAT Asn	GAG Glu	TCT Ser	ATA Ile	AAT Asn 780	AAA Lys	GCT Ala	ATG Met	ATT Ile		2352
AAT Asn 785	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 790	AAT Asn	CAA Gln	TGC. Cys	TCT Ser	GTT Val 795	TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 800		2400
			CCT Pro														2448
			GCA Ala 820														2496
			GTA Val														2544 <sup>;</sup>
			CCT Pro							Val							2592
			TTT Phe						TAA *						,		2622

### (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 874 amino acids (B) TYPE: amino acid

  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val

Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met 20 25 30

Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro 35 40 45

Glu	Arg 50	Asp	Thr	Phe	Thr	Asn 55	Pro	Glu	Glu	Gly	Asp 60	Leu	Asn	Pro	Pro
Pro 65	Glu	Ala	Lys	Gln	Val 70	Pro	Val	Ser	Tyr	Tyr 75	Asp	Ser	Thr	Tyr	Leu 80
Ser	Thr	Asp	Asn	Glu 85	Lys	Asp	Asn	Tyr	Leu 90	Lys	Gly	Val	Thr	Lys 95	Leu
Phe	Glu	Arg	Ile 100	Tyr		Thr	Asp	Leu 105	Gly	Arg	Met	Leu	Leu 110	Thr	Şer
Ile	Val	Arg 115	Gly	Ile	Pro	Phe	Trp 120	Gly	Gly	Ser	Thr	Ile 125	Asp	Thr	Glu
Leu	Lys 130	Val	Ile	Asp	Thr	Asn 135	Cys	Ile	Asn	Val	Ile 140	Gln	Pro	Asp	Gly
Ser 145	Tyr	Arg	Ser	Glu	Glu 150	Leu	Asn	Leu	Val	11e 155	Ile	Gly	Pro	Ser	Ala 160
Asp	Ile	Ile	Gln	Phe 165	Glu	Cys	Lys	Ser	Phe 170	Gly	His'	Glu	Val	Leu 175	Asn
Leu	Thr	Arg	Asn 180	Gly	Tyr	Gly	Ser	Thr 185	Gln	Tyr	Ile	Arg	Phe 190	Ser	Pro
Asp	Phe	Thr 195	Phe	Gly	Phe	Glu	Glu 200	Ser	Leu	Glu	Val	Asp 205	Thr	Asn	Pro
Leu	Leu 210	Gly	Ala	Gly	Lys	Phe 215	Ala	Thr	Asp	.Pro	Ala 220	Val	Thr	Leu	Ala
His 225	Glu	Leu	Ile	His	Ala 230	Gly	His	Arg	Leu	Tyr 235	Gly	Ile	Ala	Ile	Asn 240
Pro	Asn	Arg	Val	Phe 245	Lys	Val	Asn	Thr	Asn 250	Ala	Tyr	Tyr	Glu	Met 255	Ser
Gly	Leu	Glu	Val 260	Ser	Phe	Glu	Glu	Leu 265	Arg	Thr	Phe	Gly	Gly 270	His	Asp
Ala	Lys	Phe 275	Ile	Asp	Ser	Leu	Gln 280	Glu	Asn	Glu	Phe	Arg 285	Leu	Tyr	Tyr
Tyr	Asn 290	Lys	Phe	Lys	Asp	Ile 295	Ala	Ser	Thr	Leu	Asn 300	Lys	Ala	Lys	Ser
11e 305	Val	Gly	Thr	Thr	Ala 310	Ser	Leu	Gln	Tyr	Met 315	Lys	Asn	Val	Phe	Lys 320
Glu	Lys	Tyr	Leu	Leu 325	Ser	Glu	Asp	Thr	Ser 330	Gly	Lys-	-Phe	Ser	Val 335	Asp
Lys	Leu	Lys	Phe 340	Asp	Lys	Leu	Tyr	Lys 345	Met	Leu	Thr	Glu	Ile 350	Tyr	Thr
Glu	Asp	Asn 355	Phe	Val	Lys	Phe	Phe 360	Lys	Val	Leu	Asn	Arg 365	Lys	Thr	Tyr
Leu	Asn 370	Phe	Asp	Lys	Ala	Val 375		Lys	Ile	Asn	Ile 380	Val	Pro	Lys	Val
Asn 385	Tyr	Thr	Ile	Tyr	Asp 390	Gly	Phe	Asn	Leu	Arg 395	Asn	Thr	Asn	Leu	Ala 400

Al	a As	n I	Phe	As	n Gl 40	y Gl 5	n As	n Th	r Gl	u Il 41		n As	n Me	t Ası	n Ph 41	
Ly	s Le	u I	ys	42	n Ph O	e Th	r Gl	y Le	u Ph 42	e Gl	u Ph	e Ty	r Ly:	s Lei 430	ı Le	u Cy
Va	l Ar	g G 4	1y 35	11	e Il	e Th	r Se	r Ly:	s Th	r Ly:	s Se	r Lei	1 Ası 44!	p Lys	Gl:	у Ту
As	n Ly 45	s A O	la.	Le	u As	n Asi	p Le	u Cy:	s Ile	e Lys	s Va	1 Asr 460		n Trp	) Ası	p Le
Ph 46		e S	er	Pro	Se	r Gl:	u Ası	p Ası	n Phe	Thi	475		Leu	ı Asn	Lys	5 Gly
Gli	u Gl	u I	le	Thi	Se:	r Ası	Thi	r Asr	ıle	Glu 490	ı Alá	a Ala	Glu	. Glu	Asr 499	
Se	r Le	u A	sp	Let 500	ı Ile	e Glr	n Glr	Tyr	Ty:	Lev	Thr	Phe	Asn	Phe 510		) Asr
Gli	ı Pro	o, G. 5:	lu 15	Asr	ılle	e Ser	Ile	Glu 520	Asn	Leu	Ser	Ser	Asp 525	Ile	Ile	G G J S
Glr	530	ı G:	lu	Leu	Met	Pro	Asn 535	Ile	Glu	Arg	Phe	Pro 540		Gly	Lys	Lys
Tyr 545		Le	eu	Asp	Lys	Tyr 550	, Thr	Met	Phe	His	Tyr 555		Arg	Ala	Gln	Glu 560
Phe	Glu	Ні	s	Gly	Lys 565	Ser	Arg	Ile	Ala	Leu 570		Asn	Ser	Val	Asn 575	
Ala	Lev	Le	u	Asn 580	Pro	Ser	Arg	Val	Tyr 585	Thr	Phe	Phe	Ser	Ser 590	Asp	Tyr
Val	Lys	Ly 59		Val	Asn	Lys	Ala	Thr 600	Glu	Ala	Ala	Met	Phe 605	Leu	Gly	Trp
Val	Glu 610		n i	Leu	Val	Tyr	Asp 615	Phe	Thr	Asp	Glu	Thr 620	Ser	Glu	Val	Ser
Thr 525	Thr	As	p 1	ŗ'ns	Ile	Ala 630	Asp	Ile	Thr	Ile	Ile 635	Ile	Pro	туг	Ile	Gly 640
Pro	Ala	Le	u A	Asn	Ile 645	Gly	Asn	Met	Leu	Tyr 650	Lys	Asp	Asp	Phe	Val 655	Gly
lla	Leu	Il		Phe 560	Ser	Gly	Ala	Val	Ile 665	Leu	Leu	Glu	Phe	Ile 670	Pro	Glu
le	Ala	11 67		Pro	Val	Leu	Gly	Thr 680	Phe	Ala	Leu	Val	Ser 685	Tyr	Ile	Ala
sn	Lys 690	Va:	1 1	eu	Thr	Val	Gln 695	Thr	Ile	Asp	Asn	Ala 700	Leu	Ser	Lys	Arg
.sn 05	Glu	Ly	s T	,rp	Asp	Glu 710	Val	Tyr	Lys	Tyr	Ile 715	Val	Thr	Asn	Trp	Leu 720
la	Lys	Va:	L A	sn	Thr 725	Gln	Ile	Asp	Leu	Ile 730	Arg	Lys	Lys	Met	Lys 735	Glu
la	Leu	Glu		sn	Gln	Ala	Glu		Thr	Lys	Ala	Ile	Ile	Asn	Tyr	Gln

Tyr	Asn	Gln 755	Tyr	Thr	Glu	Glu	Glu 760		Asn	Asn	Ile	Asn 765	Phe	Asn	Ile		
Asp	Asp 770	Leu	Ser	Ser	Lys	Leu 775	Asn	Glu	Ser	Ile	Asn 780	Lys	Ala	Met	Ile		
Asn 785	Ile	Asn	Lys	Phe	Leu 790	Asn	Gln	Cys	Ser	Val 795	Ser	Tyr	Leu	Met	Asn B00		
Ser	Met	Ile	Pro	Tyr 805		'Va'l	Lys	Arg	Leu 810	Glu	Asp	Phe	Asp	'Ala 815	Ser		
Leu	Lys	Asp	Ala 820	Leu	Leu	Lys	Tyr	Ile 825	Tyr	Asp	Asn	Arg	Gly 830	Thr	Leu		
lle	Gly	Gln 835	Val	Asp	Arg	Leu	Lys 840	Asp	Lys	Val	Asn	Asn 845	Thr	Leu	Ser		
Thr	Asp 850	Ile	Pro	Phe	Gln	Leu 855	Ser	Lys	Tyr	Val	Asp 860		Gln	Arg	Leu		2
Leu 865	Ser	Thr	Phe	Thr	Glu 870	Tyr	Ile	Lys	*							i	
(2)	INFO	ORMAT	CION	FOR	SEQ	ID N	10 :	7 :		,					-		
•	(i)	( <i>I</i> (E	A) LI 3) T' 2) ST	CE CI ENGTI YPE: TRANI OPOLO	nucl	513 'l leic ESS:	ase acid doul	pai:	cs						s.		
	(ii)	MOI	LECUI	LE TY	PE:	DNA	(ger	omic	:)								
	(ix)	(7		E: AME/I DCATI			513 <sup>-</sup>										•
	(xi)	SEC	QUENC	E DE	ESCRI	PTIC	N: 9	SEQ 1	D NC	): 7:					•		
ATG Met 1	CCA Pro	TTT Phe	GTT Val	AAT Asn 5	aaa Lys	CAA Gln	TTT Phe	AAT Asn	TAT Tyr 10	AAA Lys	GAT Asp	CCT Pro	GTA Val	AAT Asn 15	GGT Gly		48
TT /al	GAT Asp	ATT Ile	GCT Ala 20	TAT Tyr	ATA Ile	AAA Lys	ATT Ile	CCA Pro 25	TAA naA	GCA Ala	GGA Gly	CAA Gln	ATG Met 30	CAA Gln	CCA Pro	•	96
STA /al	AAA Lys	GCT Ala 35	TTT Phe	AAA Lys	ATT Ile	CAT His	AAT Asn 40	AAA Lys	ATA Ile	TGG Trp	GTT Val	ATT Ile 45	CCA Pro	GAA Glu	AGA Arg		144
ASP	ACA Thr 50	TTT Phe	ACA Thr	AAT Asn	CCT Pro	GAA Glu 55	GAA Glu	GGA Gly	GAT Asp	TTA Leu	AAT Asn 60	CCA Pro	CCA Pro	CCA Pro	GAA Glu		192
GCA Ala 65	AAA Lys	CAA Gln	GTT Val	CCA Pro	GTT Val 70	TCA Ser	TAT Tyr	TAT Tyr	GAT Asp	TCA Ser 75	ACA Thr	TAT Tyr	TTA Leu	AGT Ser	ACA Thr 80		240
TAZ qz	AAT Asn	GAA Glu	AAA Lys	GAT Asp 85	AAT Asn	TAT Tyr	TTA Leu	AAG Lys	GGA Gly 90	GTT Val	ACA Thr	AAA Lys	TTA Leu	TTT Phe 95	GAG Glu		288

				Thi					Met					Ile	GTA Val	336
			Pro					Ser					Glu		AAA Lys	384
		Asp					Asn					Asp			TAT	432
AGA Arg 145	Ser	GAA Glu	GAA Glu	CTT Leu	AAT Asn 150	Leu	GTA Val	ATA	ATA Ile	GGA Gly 155	CCC Pro	TCA Ser	GCT Ala	GAT Asp	ATT Ile 160	480
ATA Ile	CAG Gln	TTT Phe	GAA Glu	TGT Cys 165	Lys	AGC Ser	TTT Phe	GGA Gly	CAT His	GAA Glu	GTT Val	TTG Leu	AAT Asn	CTT Leu 175	ACG Thr	528
			TAT Tyr 180													576
ACA Thr	TTT Phe	GGT Gly 195	TTT Phe	GAG Glu	GAG Glu	TCA Ser	CTT Leu 200	GAA Glu	GTT Val	GAT Asp	ACA Thr	AAT Asn 205	CCT Pro	CTT Leu	TTA Leu	624
			AAA Lys													672
			GCT Ala													720
			AAA Lys													768
			TTT Phe 260													B16.
			AGT Ser													. 864
			GAT Asp													912
			GCT Ala													960
			TCT Ser													1008
		Asp	AAG Lys 340				Met									1056
AAT Asn	Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104

TTT Phe	GAT Asp 370	Lys	GCC	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr		1152
ACA Thr 385	Ile	TAT Tyr	GAT Asp	GGA Gly	Phe 390	Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400		1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT	ACT	AAA Lys 415	CTA Leu		1248
					TTG Leu												1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys		1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	'GAC Asp	TTG Leu	TTT Phe	TTT Phe	,	1392
					AAT Asn 470												1440
					AAT Asn												1488
					TAT Tyr												1536
					GAA Glu												1584
					ATA Ile												1632
					ATG Met 550											1	1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu		1728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys		1776
AAA Lys					ACG Thr												1824
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr		1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640		1920

					Met					Asp					TTA Leu	•	1968
ATA Ile	TTT Phe	TC# Ser	GG/ Gly 660	/ Ala	GTI Val	ATT Ile	Leu	Leu 665	Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT	GCA Ala		2016
ATA Ile	CCT	GTA Val 675	Leu	GGT Gly	ACI Thr	TTT Phe	GCA Ala 680	Leu	GTA Val	TCA Ser	TAT Tyr	Ile 685	GCG Ala	AAT Asn	AAG Lys		2064
GTT Val	CTA Leu 690	ACC	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	Lys	AGA Arg	AAT Asn	GAA Glu		2112
AAA Lys 705	Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720		2160
			CAG Gln														2208
			GCA Ala 740														2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp		2304
			AAA Lys														2352
			TTG Leu														2400
			GGT Gly														2448
			TTA Leu 820														2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp		2544
			CAG Gln		Ser												2592
			GAA Glu	Tyr													2613

(2) INFORMATION FOR SEQ ID NO: 8:

<sup>(</sup>i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 871 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

1 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu . 325 330 335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 360 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 470 Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu. Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 505 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 535 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 550 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 600 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys

												•			1	
Val	Leu 690	Thr	Val	Gln	Thr	Ile 695	Asp	Asn	Ala	Leu	Ser 700	Lys	Arg	Asn	Glu	
Lys 705	Trp	Asp	Glu	Val	Tyr 710	Lys	Tyr	Ile	Val	Thr 715	Asn	Trp	Leu	Ala	Lys 720	
Val	Asn	Thr	Gln	Ile 725	Asp	Leu	Ile	Arg	Lys 730	Lys	Met	Lys	Glu	Ala 735	Leu	
Glu	Asn	Gln	Ala 740	Glu	Ala	Thr	Lys	Ala 745	Ile	Ile	Asn	Tyr	Gln 750	Tyr	Asn	
Gln	Tyr	Thr 755	Glu	Glu	Glu	Lys	Asn 760	Asn	Ile	Asn	Phe	Asn 765	Ile	Asp	Asp	
Leu	Ser 770	Ser	Lys	Leu	Asn	Glu 775	Ser	Ile	Asn	Lys	Ala 780	Met	Ile	Asn	Ile	
Asn 785	Lys	Phe	Leu	Asn	Gln 790	'Cys	Ser	Val	Ser	Tyr 795	Leu	Met	Asn	Ser	Met 800	*
Ile	Pro	Tyr	Gly	Val 805	Lys	Arg	Leu	Glu	Asp 810	Phe	Asp	Ala	Ser	Leu 815	Lys	
Asp	Ala	Leu	Leu 820	Lys	Tyr	Ile	Tyr	Asp 825	Asn	Arg	Gly	Thr	Leu 830	Ile	Gly	
Gln	Val	Asp 835	Arg	Leu	Lys	Asp	Lys 840	Val	Asn	Asn	Thr	Leu 845	Ser	Thr	Asp	
Ile	Pro 850	Phe	Gln	Leu	Ser	Lys 8 <b>5</b> 5	Tyr	Vaĺ	Asp	Asn	Gln 860	Arg	Leu	Leu	Ser	
Thr 865	Phe	Thr	Glu	Tyr	Ile 870	Lys										
(2)	INFO	RMAI	NOI	FOR	SEQ	ID N	10: 9	<b>)</b> :							-	
	(i)	() ()	QUENC A) LE B) TY C) SI O) TC	NGTH PE: RANI	i: 26 nucl EDNE	28 t eic SS:	ase acid	pair i	S						ı	
	(ii)	MOI	ECUL	E TY	PE:	DNA	(ger	omic	=)							
	(ix)	(P	ATURE A) NA B) LC	ME/F			28									
	(xi)	SEC	UENC	E DE	SCRI	PTIC	)N: S	SEQ 1	D NO	): 9:	:					
ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	48
GTT Val	GAC Asp	ATT Ile	GCC Ala 20	TAC Tyr	ATC Ile	AAA Lys	ATT Ile	CCA Pro 25	AAC Asn	GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
GTG Val	AAG Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT Ile	CAT His	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192

GCA Ala 65	Lys	CAC Glr	GTG Val	CCA Pro	GTI Val	Ser	TAC	TAC	GAT Asp	TCA Ser 75	Thr	TAT	CTG Leu	AGC Ser	ACA Thr 80	240
GAC Asp	AAC Asn	GAG Glu	AAC Lys	GAT Asp 85	Asn	TAC	CTG Leu	Lys	GGA Gly 90	Val	ACC	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu	288
CGT Arg	ATT Ile	TAT	Ser 100	Thr	GAC Asp	CTG Leu	GGC	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC	TCA Ser 110	ATC Ile	GTC Val	336
CGC Arg	GGA Gly	ATC Ile 115	Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC	ATT	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	. 384
GTT Val	ATT Ile 130	Asp	ACT	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GÄA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	ĠAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	.CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008

A L	AA ys	TTT Phe	GA? Asp	T AAG D Ly: 340	s re	A TA u Ty	C AA r Ly	A AT s Me	G TT. t Les 34	u Th	A GAG	G AT	T TAG	C AC.	r Gl	G GAT u Asp	1056
A A	AT	TTT Phe	GTT Val 355	LLYS	G TT	T TT e Ph	T AA e Ly	A GT. s Va 36	I Le	T AA( L Asi	C AGA	A AA g Ly:	A ACA	Ty:	T TT	G AAT u Asn	1104
T' Pl	ne A	GAT Asp 370	AAA Lys	GCC Ala	C. GT/ a Val	A TT	T AA( ≥ Ly: 37:	s Ile	A AA1 2 Asi	T ATA	A GTA E Val	CC: Pro 380	Lys	GTA Val	A AA:	TAC n Tyr	1152
T	ca i nr 1 85	ATA Ile	TAT Tyr	GAT Asp	GG/ Gly	Y Phe 390	Ası	r TT/ Let	A AGA	AAl Asr	TACA Thr 395	Ası	TTA Leu	GCA Ala	GCA Ala	A AAC A Asn 400	1200
TT Pl	T I	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	Thi	GAZ Glu	ATT	TAA 1 Raa 9	AAT Asn 410	Met	AA1 Asn	TTT Phe	ACT Thr	Lys 415	CTA Leu	1248
A.F Ly	A A	TA. LSN	TTT Phe	ACT Thr 420	Gly	Leu	TTT Phe	GAA Glu	TTT Phe 425	Tyr	AAG Lys	TTG	CTA Leu	TGT Cys 430	Val	AGA Arg	1296
GG G1	G A y I	le	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	Lys 440	Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC	AAT Asn	AAG Lys	1344
AG Se	r A	CT la 50	GAT Asp	GGG Gly	GCA Ala	TTA Leu	AAT Asn 455	GAT Asp	TTA Leu	TGT Cys	ATC Ile	AAA Lys 460	GTT Val	AAT Asn	AAT Asn	TGG	1392
GA As 46	рL	TG eu	TTT Phe	TTT Phe	AGT Ser	CCT Pro 470	TCA Ser	GAA Glu	GAT Asp	AAT Asn	TTT Phe 475	ACT Thr	AAT Asn	GAT Asp	CTA Leu	AAT Asn 480	1440
AA. Ly:	A G B G	GA (	GAA Glu	GAA Glu	ATT Ile 485	ACA Thr	TCT Ser	GAT Asp	ACT Thr	AAT Asn 490	ATA Ile	GAA Glu	GCA Ala	GCA Ala	GAA Glu 495	GAA Glu	1488
AA: Ası	r A' n I	rr i	AGT Ser	TTA Leu 500	GAT Asp	TTA Leu	ATA Ile	CAA Gln	CAA Gln 505	TAT Tyr	TAT Tyr	TTA Leu	ACC Thr	TTT Phe 510	AAT Asn	TTT Phe	1536
GA: Asi	A T	sn (	GAA Glu 515	CCT Pro	GAA Glu	AAT Asn	ATT Ile	TCA Ser 520	ATA Ile	GAA Glu	AAT Asn	CTT Leu	TCA Ser 525	AGT Ser	GAC Asp	ATT Ile	1584
ATA	: G)	SC C	CAA Sln	TTA Leu	GAA Glu	CTT Leu	ATG Met 535	CCT Pro	TAA Asn	ATA Ile	GAA Glu	AGA Arg 540	TTT Phe	CCT Pro	AAT Asn	GGA Gly	1632
AAA Lys 545	: Ly	AG I	TAT (	GAG (	Leu	GAT Asp 550	AAA Lys	TAT Tyr	ACT Thr	ATG Met	TTC Phe 555	CAT His	TAT Tyr	CTT Leu	CGT Arg	GCT Ala 560	1680
CAP Glr	GF GI	u F	TT ( he (	Glu	CAT His 565	GGT Gly	AAA Lys	TCT Ser	AGG Arg	ATT Ile 570	GCT Ala	TTA Leu	ACA Thr	AAT Asn	TCT Ser 575	GTT Val	1728
AAC	GA Gl	A G	la 1	TTA 1 Leu 1 580	TTA . Leu .	AAT Asn	CCT Pro	Ser	CGT Arg 585	GTT Val	TAT Tyr	ACA Thr	TTT Phe	TTT Phe 590	TCT Ser	TCA Ser	1776
GAC Asp	TA Ty	r V	TA 1 al 1 95	AAG 1 Lys 1	AAA ( Lys '	GTT .	Asn	AAA Lys 600	GCT . Ala	ACG Thr	GAG (	GCA Ala	GCT Ala 605	ATG Met	TTT Phe	TTA Leu	1824

GT Va 62 AT III GT VA AT III AAA Lys 705 TEG TET TAT	TA AC	TP V	al G	lu G. CG GA	in Le	u Va:	l Tyr	GA'	r TT	T AC	C GA: r Ası 620	Gl	A AC	r AG	C GAA r Glu	1872
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	ll Se 5 'A GG	r Ti	or A	CG GA	AA TA						02.	•				,
GT Va CCC Pr ATTIL Lys TGC Trr	A GG e Gl	A CC			63	s Ile	r GCC e Ala	GA:	r ATA	A ACT Thi 635	r Ile	A AT	T AT:	CC.	A TAT O Tyr 640	1920
CCC Pri ATT III		уР	CT GO	CT TI la Le 64	u Ası	r ATA	GGT Gly	' AA1 ' Asi	ATC Met 650	Let	A TAT 1 Tyr	Lys	A GAT S Asp	GA: Asi 65!	T TTT Phe	1968
AAAAA Lys	A GG 1 Gl	T GC y Al	T T1.a Le	eu Il	A TTT e Phe	TCA Ser	GGA Gly	GCT Ala 665	Va]	T ATT	CTC Leu	Leu	GAA Glu 670	Phe	T ATA	2016
AAAA Lys	A GA	G AT u Il 67	e Al	A AT a Il	A CCT e Pro	GTA Val	TTA Leu 680	GGT Gly	ACI Thr	TTI Phe	GCA Ala	CTI Leu 685	Val	TCA Ser	TAT	2064
Lys 705 TGC Trp AAA Lys	T GCC e Ala 690	a As	T AA n Ly	G GT s Va	T CTA l Leu	ACC Thr 695	GTT Val	CAA Gln	ACA Thr	ATA Ile	GAT Asp 700	AAT Asn	GCT Ala	TTA Leu	AGT Ser	2112
Tri AAA Lys	s Arg	A AA g As	T GA n Gl	A AA u Ly:	A TGG 5 Trp 710	Asp	GAG Glu	GTC Val	TAT Tyr	AAA Lys 715	Tyr	ATA Ile	GTA Val	ACA Thr	AAT Asn 720	2160
Lys	TTA Leu	A GC	A AA a Ly	G GTT s Val 725	, AAT L Asn	ACA Thr	CAG Gln	ATT Ile	GAT Asp 730	CTA Leu	ATA Ile	AGA Arg	AAA Lys	AAA Lys 735	Met	2208
TAT Tyr	A GAP	A GC:	T TT a Lei 740	ı Glu	AAT Asn	CAA Gln	GCA Ala	GAA Glu 745	GCA Ala	ACA Thr	AAG Lys	GCT Ala	ATA Ile 750	ATA Ile	AAC Asn	2256
	CAG Gln	TAT Tyr 755	Ası	CAA Gln	TAT	ACT Thr	GAG Glu 760	GAA Glu	GAG Glu	AAA Lys	AAT Asn	AAT Asn 765	ATT Ile	AAT Asn	TTT Phe	2304
AAT Asn	Ile 770	Asp	GAT Asp	TTA Leu	AGT Ser	TCG Ser 775	AAA Lys	CTT Leu	AAT Asn	GAG Glu	TCT Ser 780	ATA Ile	AAT Asn	AAA Lys	GCT Ala	2352
ATG Met 785	ATT Ile	AAT Asn	ATA Ile	AAT Asn	AAA Lys 790	TTT Phe	TTG . Leu .	AAT Asn	CAA Gln	TGC Cys 795	TCT Ser	GTT Val	TCA Ser	TAT Tyr	TTA Leu 800	. 2400
ATG Met	AAT Asn	TCT Ser	ATG Met	ATC Ile 805	CCT Pro	TAT Tyr	GGT (	Val	AAA Lys 810	CGG Arg	TTA Leu	GAA Glu	GAT Asp	TTT Phe 815	GAT Asp	2448
GCT Ala	AGT Ser	CTT Leu	AAA Lys 820	Asp	GCA Ala	TTA Leu	Leu :	AAG Lys 825	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp	TAA nsa 830	AGA Arg	GGA Gly	2496
ACT Thr	TTA Leu	ATT Ile 835	GGT Gly	CAA Gln	GTA Val	Asp 3	AGA : Arg 1 840	ITA Leu	AAA Lys	GAT Asp	Lys	GTT Val 845	AAT Asn	AAT Asn	ACA Thr	2544
CTT Leu	AGT Ser 850	ACA Thr	GAT Asp	ATA Ile	CCT Pro	TTT ( Phe ( 855	CAG ( Gln I	CTT Leu	TCC . Ser	Lys	TAC Tyr 860	GTA Val	GAT Asp	AAT Asn	CAA Gln	2592
					TTT A				Ile :		TAA					2628

- (2) INFORMATION FOR SEQ ID NO: 10:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 876 amino acids
    - (B) TYPE: amino acid
    - (D) TOPOLOGY: linear

  - (ii) MOLECULE TYPE: protein
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 150

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 200

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
420 425 430

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435
440
445

Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp
450 460

Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn 465 470 475 480

Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu 485 490 495

Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe 500 510

Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile 515 520 525

Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly 530 535 540

Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala 545 550 555 560

Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val 565 570 575

Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe-Phe Ser Ser 580 585 590

Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu 595 600 605

Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu 610 620

Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr 625 630 635 640

Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe 645 650 655

Va]	. Gly	Ala	660	Ile	e Phe	Ser	Gly	Ala 665	Va]	Ile	Leu	Leu	670		lle		
Pro	Glu	1le 675	Ala	Ile	Pro	Val	Leu 680	Gly	Thr	Phe	: Ala	Leu 685		Ser	Tyr	÷	
Ile	Ala 690	Asn	Lys	Val	. Leu	Thr 695	Val	Gln	Thr	Ile	Asp 700		Ala	Leu	Ser		
Lys 705	Arg	Asn	Glu	Lys	Trp 710	Asp	Glu	Val	Tyr	Lys 715		Ile	Val	Thr	Asn 720		
Trp	Leu	Ala	Lys	Val 725	Asn	Thr	Gln	Ile	Asp 730		Ile	Arg	Lys	Lys 735	Met		
Lys	Glu	Ala	Leu 740	Glu	Asn	Gln	Ala	Glu 745	Ala	Thr	Lys	Ala	Ile 750		Asn		
Tyr	Gln	Tyr 755	Asn	Gln	Tyr	Thr	Glu 760	Glu	Glu	Lys	Asn	Asn 765	Ile	Asn	Phe		
Asn	Ile 770	Asp	Asp	Leu	Ser	Ser 775	Lys	Leu	Asn	Glu	Ser 780	Ile	Asn	Lys	Ala		
Met 785	Ile	Asn	Ile	Asn	Lys 790	Phe	Leu	Asn	Gln	Cys 795	Ser	Val	Ser	туг	Leu 800		
Met	Asn	Ser	Met	Ile 805	Pro	Tyr	Gly	Val	Lys 810	Arg	Leu	Glu	Asp	Phe 815	Asp		
Ala	Ser	Leu	Lys 820	Asp	Ala	Leu	Leu	Lys 825	Tyr	Ile	Tyr	Asp	Asn 830	Arg	Gly	•	
Thr	Leu	Ile 835	Gly	Gln	Val	Asp ——	Arg 840	Leu	Lys	Asp	Lys	Val 845	Asn	Asn	Thr		
Leu	Ser 850	Thr	Asp	Ile	Pro	Phe 855	Gln	Leu	Ser	Lys	Tyr 860	Val	Asp	Asn	Gln		
Arg 865	Leu	Leu	Ser '	Thr	Phe 870	Thr	Glu	Tyr	Ile	Lys 875	•	•					
.(2)	INFO	RMAT	ION :	FOR	SEQ	ID N	0: 1	1:								•	
	(i)	A) B) (C	UENCI ) LEI ) TYI ) STI ) TOI	ngth Pe : Rand	: 26 nucl EDNE	37 b eic SS:	ase acid doub	pair	s ,							•	
	(ii)	MOL	ECULI	TY	PE:	DNA	(gen	omic	)								
	(ix)	(A)	TURE:	Æ/K			37										
	(xi)	SEQ	JENCE	E DE	SCRI	PTIO	N: S	EQ I	D NO	: 11	:						
ATG Met 1	CAG : Gln	TTC ( Phe V	STG A Val A	AC Asn S	AAG ( Lys (	CAG '	Phe	AAC ' Asn '	TAT Tyr 10	AAG Lys	GAC Asp	CCT (	GTA Val	AAC Asn 15	GGT Gly	4	ξ

GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

BNSDOCID: <WO\_\_\_9807864A1\_I\_>

GT Va	G AA 1 Ly	s Al	T Ti a Pi	CC AA ne Ly	G AT	T CAT	T AA( S Asi 4(	Ly:	A ATO	TGC Tr	G GT	r AT	e Pro	G GA	A CGC u Arg	144
GA'	T AC	r Ph	T AC	G AA Ir As	c cc n Pr	G GAA O Glu 55	ı Glu	GG/ Gly	A GAO / Asp	TTC Let	AA E 12A 1	n Pro	CCC Pro	CCO Pro	G GAA	192
GC/ Ala 6	a Ly	G CA s Gl	G G1 n Vä	G CC	A GT o Va 7	l Ser	TAC Tyr	TAC	GAT Asp	TCA Ser 75	Thr	TAT	CTC	AG(	C ACA Thr 80	240
GA( Asi	C AA( p Asi	G GA	G AA u Ly	G GA S As 8	p Ası	TAC Tyr	CTG Leu	Lys	GGA Gly 90	' Val	ACC Thr	Lys	TT#	TTO Phe 95	GAG Glu	288
CGT	T ATT	TA Ty	T TC r Se 10	C AC' r Th: 0	T GAO	C CTG	GGC	CGT Arg 105	Met	CTG Leu	CTG Leu	ACC	Ser 110	Ile	GTC Val	336
'CGC	GGA Gly	ATC 110	e Pr	A TT	r TGC ≥ Trp	G GGT	GGC Gly 120	AGT Ser	ACC	ATT	GAC Asp	ACG Thr 125	Glu	Leu	AAG Lys	384
GT1 Val	Ile 130	Ası	C AC	r Aac	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC	TAC Tyr	432
AGA Arg 145	Ser	GA/ Glu	GAI Glu	A CTI 1 Leu	AAC Asn 150	Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTI Phe	GAC Glu	TGC Cys 165	Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC	TCT	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	. 672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
				GTT Val 245												768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	Arg '	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
				TTG Leu		Glu .					Leu					864
				ATT Ile	Ala					Lys .						912

GG1 G1 <sub>3</sub> 305	Thi	ACT Thi	GCT Ala	TCA Ser	Leu 310	ı Gln	TAT	ATC Met	AAA Lys	AAT Asn 315	Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Sex	GAA Glu 325	Asp	ACA Thr	TCT	GGA Gly	AAA Lys 330	Phe	TCG Ser	GTA ,Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT	GAT Asp	AAC Lys 340	Leu	TAC	AAA Lys	ATG Met	Leu 345	Thr	GAG Glu	ATT	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTI Val 355	Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	Lys	GCC Ala	GTA Val	TTT	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	Phe	TAT Tyr	Lys	TTG Leu	Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA -Ile	ATA Ile 435	ACT Thr	TCT	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
ATC	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
GAT Asp	CTA Leu	TAA Asn	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	GAT Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
			GAT Asp			Pro										1584
			ATA Ile		Gln											1632
			AAA Lys	Lys										His		1680
			CAA Gln													1728

AA7 Asr	TC Se	T GI r Va	T AA 1 As 58	n Gl	AA GC	CA TI	A TT u Le	A AA u As 58	n Pr	T AG o Se:	r Ar	T GT	T TA' 1 Ty: 59	r Th	A TTT r Phe	177	76
TTT Phe	TC' Se	T TC r Se 59	r As	C TA p Ty	AT GI Yr Va	A AA l Ly	G AA s Ly 60	s Va	T AA 1 As:	T AAJ n Lys	A GC	ACC Thi	r Glı	G GC	A GCT a Ala	182	24
ATG Met	Phe 610	e Le	A GG u Gl	C TG y Tr	G:GT p Va	A GA 1 G1: 61:	u Gl	A TT.	A GT	A TAT	GAT Asp 620	Phe	r ACC	GAT Asp	GAA Glu	. 187	2
ACT Thr 625	Ser	GA.	A GT. u Va	A AG l Se	T AC r Th 63	r Th	G GA: r As	T AAA	A AT	GCC Ala 635	Asp	T ATA	ACT Thi	TATA Ile	ATT : Ile 640	192	0
ATT Ile	Pro	TA' Ty:	r AT	A GG. ∋ G1: 64:	y Pr	T GCT o Ala	TTI Lev	AA A 1 Asi	ATA 1 116 650	e Gly	AAT Asn	ATG Met	Leu	TAT Tyr 655	Lys	196	8
GAT Asp	GAT Asp	TTT Phe	GTA Val 660	Gly	r GC' y Ala	T TTA a Lei	ATA 11e	Phe 665	e Ser	GGA Gly	GCT Ala	GTT Val	Ile 670	Leu	TTA Leu	201	6
GAA Glu	TTT	ATA Ile 675	Pro	GAC Glu	3 AT.	Γ GCA ∋ Ala	ATA Ile 680	Pro	GTA Val	TTA Leu	GGT	ACT Thr 685	TTT Phe	GCA Ala	CTT Leu	206	4
GTA Val	TCA Ser 690	TAT	ATT	GCC Ala	AAT AST	AAG Lys 695	GTI Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	ACA Thr	ATA Ile	GAT Asp	AAT Asn	2112	2
GCT Ala 705	TTA Leu	AGT Ser	AAA Lys	AGA Arg	AAT ASN 710	Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 715	GTC Val	TAT Tyr	AAA Lys	TAT	ATA Ile 720	2160	)
GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 725	Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg	2208	3
AAA Lys	AAA Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala	2256	5
ATA Ile	ATA Ile	AAC Asn 755	TAT Tyr	CAG Gln	TAT	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AAT , Asn	2304	<b>L</b>
ATT .	AAT Asn 770	TTT Phe	AAT Asn	ATT Ile	GAT Asp	GAT Asp 775	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	AAT Asn	GAG Glu	TCT Ser	ATA Ile	2352	!
AAT . Asn 785	AAA Lys	GCT Ala	ATG Met	ATT Ile	AAT Asn 790	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 795	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 800	2400	)
TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GAA Glu	2448	
GAT 1	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT Tyr	GAT Asp	2496	
AAT A	Arg															2544	

- 64 -

AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA
Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
850

GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA
Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys \*

- (2) INFORMATION FOR SEQ ID NO: 12:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 879 amino acids
    - (B) TYPE: amino acid
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 .200 .205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 300 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 345 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 360 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 470 475 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 490 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 520 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 535 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 570 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 585 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu

Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 630

Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 650

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 680

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 695

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 730

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 745

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 855

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 870

- (2) INFORMATION FOR SEQ ID NO: 13:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2862 base pairs

    - (B) TYPE: nucleic acid(C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
  - (ix) FEATURE:
    - (A) NAME/KEY: CDS
    - (B) LOCATION:1..2862
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Me	G CA t Gl	G T	rc ( ne '	GTG Val	As	C AA n Ly 5	G CA	AG T	TC A	AAC Asn	TA1	Ly	G GA s As	C CC p Pr	T G	TA al	AAC Asi	Gl	T Y	48
GT: Va	r ga l As	C AT	TT (	GCC Ala 20	Ty	C AT	C AA e Ly	A A	TT C	CA Pro 25	AAC Asr	GC Ala	C GG a Gl	C CA y Gl	n M	TG et 30	CAC Glr	G CC	G 5	96
GTC Val	Ly:	s Al	T 1	TC Phe	AA( Lys	G AT	T CA e'Hi	s As	AC A sn L	AA ys	ATC	TG	G GT	T AT	T C( e P: 5	CG ro	GA# Glu	CG(	3	. 144
GAT Asp	ACA Thi	r Ph	T A	CG hr	AAC Asr	CCC Pro	G GA o Gl 5	u Gl	va G .u G	GA <sup>.</sup> ly	GAC Asp	TT(	AA AS:	C CC n Pr 0	G CC	CG CO	Pro	GAA Glu	1	192
GCA Ala 65	Lys	G CA G Gl	G G n V	TG	CCA	GT: Vail 70	l Se	A TA r Ty	C T	AC yr	GAT Asp	TCA Ser 75	Th	TA Ty	r Cr	rG eu	AGC Ser	ACA Thr		240
GAC Asp	AAC Asn	GA G1	G A u L	AG ys	GAT Asp 85	AAC Asr	TAC Ty:	C CT	G A	AG YS	GGA Gly 90	GTG Val	AĆ(	Ly:	A TI	A u	TTC Phe 95	GAG Glu	<b>;</b>	288
CGT Arg	ATT	TA'	r Se	cc er 00	ACT Thr	GAC Asp	CTC Lev	G GG	C CC y Ai	rg	ATG Met	CTG Leu	Lei	ACO Th:	T TC F Se 11	r	ATC Ile	GTC Val		336
CGC Arg	GGA Gly	116 119	2 P1	CA ro	TTT Phe	TGG	GG1 Gly	GG( Gl)	y Se	T	ACC Thr	ATT Ile	GAC Asp	ACC Thi	Gl	G u	TTG Lẹu	AAG Lys		384
GTT Val	ATT Ile 130	Asp	AC Th	r .	AAC Asn	TGC Cys	ATT Ile 135	Ası	C GT n Va	G.	ATC Ile	CAA Gln	CCA Pro	GAC Asp	GG G1	T ;	AGC Ser	TAC Tyr		432
AGA Arg 145	TCT Ser	GAP Glu	GA G1	u i	CTT Leu	AAC Asn 150	CTC	GT# Val	A AT	c :	ATC Ile	GGG Gly 155	CCC	TCC	GC Al	G (	Asp	ATT Ile 160		480
ATC Ile	CAG Gln	TTT Phe	GA G1	u (	IGC Cys 165	AAG Lys	AGC Ser	TT1 Phe	GG Gl	y I	CAC His 170	GAA Glu	GTG Val	TTG	AA:	n I	CTG Leu 175	ACG Thr		528
CGT Arg	AAC Asn	GGT Gly	TA Ty 18	r C	GGC Sly	TCT Ser	ACT Thr	CAG Gln	TA:	r l	ATT Ile	CGT Arg	TTC Phe	AGC Ser	Pro	o A	GAC Asp	TTC Phe	•	576
ACG Thr	TTC Phe	GGT Gly 195	TT(	e G	GAG Slu	GAG Glu	AGC Ser	CTG Leu 200	Gl	G C	GTT /al	GAT Asp	ACC Thr	AAC Asn 205	Pro	3 C	TG Leu	TTG Leu		624
ily .	GCA Ala 210	GGC Gly	AA( Lys	G T	TC he	Ala	ACT Thr 215	GAT Asp	Pro	A G	SCG (	GTG Val	ACC Thr 220	CTG Leu	GC# Ala	A C	CAC	GAG Glu		672
TG i eu i 25	ATC Ile	CAC His	GCC	C G	ly I	TAT eiH 230	CGT Arg	CTG Leu	TAT	G	ly :	ATT Ile 235	GCG Ala	ATT Ile	AAC Asr	C C	ro	AAC Asn 240		720
GC (	GTG Val	TTC Phe	AAC Lys	; V	TT / al / 45	AAC Asn	ACC Thr	AAC Asn	GCC	T	AC : yr : 50	TAC Tyr	GAG Glu	ATG Met	AGT Ser	G	GT 1y 55	TTA Leu		768
AA C lu V	GTA /	AGC Ser	TTC Phe 260	G.	AG C	SAA ( Slu )	CTG Leu	CGC Arg	ACG Thr 265	P	TC (	GGT Gly	GGC Gly	CAT His	GAT Asp 270	A	CG . la .	AAG Lys		816

TT:	T ATC	GAC Asp 275	Se	C TTO	G CAG	G GAO	AA( 1 Ast 28(	ı Glı	TTC 1 Phe	C CGT	CTC Lev	TAC Tyr 285	Tyr	TAC Tyr	AAC Asn	<sub>.</sub> 864
AAC Lys	TTT Phe 290	Lys	GA:	r AT	C GC/ E Ala	A AGT a Ser 295	Thr	CTC Lev	AAC Asr	AAC Lys	GCT Ala 300	Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	Thr	ACT	GCT Ala	TC/ Sei	TTA Leu 310	Glr	TAT Tyr	ATC Met	AAA Lys	AAT Asn 315	Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT	CTC	CTA Leu	TCI Ser	GAA Glu 325	ı Asp	ACA Thr	TCT Ser	GGA Gly	Lys 330	Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	Leu	TAC	AAA Lys	ATG Met	TTA Leu 345	Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTI Phe	TTT Phe	AAA Lys	GTA Val 360	Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT	TTG Leu	AAT Asn	1104
TTT	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	,1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gl <sup>i</sup> y	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
ATC Ile	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	GAT Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
						CCT Pro										1584
					Gln	TTA Leu 535										1632

CCT Pro 545	Asn	GG G1	SA AJ .y Ly	AA AU /s L	s T	AT G/ /r G/ 50	AG TI	TA GA eu As	AT AU Sp Ly	AA TA /s Ty 55	r Tr	T AT	G TI	CC CA	AT TAT is Tyr 560	
CTT Leu	CGT Arg	GC	T CA a Gl	A GA .n G] 56	u Pr	TT GA ne Gl	VA CA Lu Hi	AT GO	T A# .y Ly 57	's Se	T AC	G AT	T GC e Al	T TT a Le 57	TA ACA tu Thr	1728
AAT Asn	TCT Ser	GT Va	T AA l As 58	n Gi	A GC u Al	A TI a Le	A TT	A AA u As 58	n Pr	T AG	T CG r Ąr	T GT g Va	T TA 1 Ty 59	r Th	A TTT r Phe	, 1776
TTT Phe	TCT Ser	TC. Se: 59:	r As	C TA p Ty	T GT r Va	A AA 1 Ly	G AA s Ly 60	s Va	T AA l As	T AA n Ly	A GC s Al	T AC a Th	r Gl	G GC u Al	A GCT a Ala	1824
ATG Met	TTT Phe 610	TT	A GG u Gl	C TG	G GT p Va	A GA 1 Gl 61	u Gl	A TT. n Le	A GT u Va	A TA' 1 Ty:	T GA r As 62	p Pho	r ac	C GA r As	T GAA p Glu	1872
ACT Thr 625	AGC, Ser	GA/ Glu	A GTA	A AGʻ l Se:	r AC r Th: 63	r Th	G GA' r As <sub>l</sub>	T AA	A AT	T GCC e Ala 639	a As	T ATA	A AC	r AT.	A ATT e Ile 640	1920
ATT Ile	CCA Pro	TAI Tyr	T ATA	GG/ Gly 645	/ Pro	r GC: o Ala	TTI a Let	A AA? 1 Asi	1 AT	e Gly	AA 1 Asi	T ATO	TT/	A TA	T AAA r Lys	1968
GAT Asp	GAT Asp	TTT Phe	GTA Val 660	. Gly	GCT Ala	TTA Leu	ATA 1 Ile	TTT Phe	: Sei	A GGA c Gly	GCT Ala	r GT7 a Val	116 670	Let	TTA Leu	2016
GAA Glu	Phe	ATA Ile 675	CCA Pro	GAG Glu	ATT Ile	GCA Ala	ATA Ile 680	Pro	GT#	Leu	GGT Gly	Thr 685	Phe	GCA Ala	CTT Leu	2064
Val :	TCA Ser	TAT Tyr	ATT Ile	GCG Ala	AAT Asn	Lys 695	Val	CTA Leu	ACC	GTT Val	CAA Glm 700	Thr	ATA	GAT Asp	TAAT Asn	2112
GCT : Ala I 705	rta / Leu S	AGT Ser	AAA Lys	AGA Arg	AAT Asn 710	GAA Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 715	GTC Val	TAT Tyr	AAA Lys	TAT	ATA Ile 720	2160
GTA A	ACA / Thr /	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA, Arg	2208
AAA A Lys I	AA A ys M	ATG let	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala	2256
ATA A Ile I	le A	AC sn 55	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AAT Asn	2304
ATT A Ile A 7	AT T sn P 70	TT he	AAT Asn	ATT Ile	GAT Asp	GAT Asp 775	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	AAT Asn	GAG Glu	TCT Ser	ATA Ile	2352
AAT A Asn L 785	AA G ys A	CT I	ATG Met	Ile	AAT Asn 790	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 795	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 800	2400
TCA T	AT T	TA j eu M	Met .	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GAA Glu	2448

- 70 -

GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	Tyr	ATA Ile 830	TAT Tyr	GAT Asp		2496
AAT Asn	AGA Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT Ile	GGT Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val		2544
AAT Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC	AAA Lys	TAC Tyr	GTA Val		2592
			AGA Arg														2640
			GAG Glu														2688
			GGA Gly 900												TAT , Tyr		2736
			AGT Ser														2784
			ÄGC Ser														2832
			GCC Ala								, <u></u>					:	2862

#### (2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 954 amino acids
  - (B) TYPE: amino acid(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 135 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys'Lys'Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 230 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 360 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val

As:	n As 5	n T	rp A	sp	Let	47	e Ph O	e Se	r Pi	o ș	er	Glu 475	Asp	Ası	n Phe	= Th:	Asn 480
Asj	p Le	u As	sn L	ys	Gl <sub>y</sub> 485	Gl:	ı Gl	u Il	e Th	r S	er 90'	Asp	Thr	Ası	ılle	Glu 499	ı Ala
Ala	a Gl	u Gi	lu A 5	sn 00	Ile			u As	p Le 50	u I. 5	le	Gln	Gln	Туз	ту: 510	Lei	Thr
Ph€	Ası	n Pł 51	ne A: .5	sp	Asn	Gli	ı Pro	520	u As O	n Il	le	Ser	Ile	Glu 525		Leu	Ser
Ser	530	o Iʻl	e I	le	Gly	Glr	Let 535	u Gli	ı Le	u Me	t	Pro	Asn 540	Ile	Glu	Arg	' Phe
Pro 545	Asr	Gl	y Ly	/S	Lys	Tyr 550	Gli	ı Lev	ı As	p Ly	'S	Tyr 555	Thr	Met	Phe	His	Tyr 560
Leu	Arg	, Al	a Gl	n.	Glu 565	Phe	Glu	1 His	G1	y Ly 57	s : 0	Ser	Arg	,Ile	Ala	Leu 575	Thr
Asn	Ser	· Va	1 As 58	0	Glu	Ala	Leu	Leu	1 Ası 58	n Pr	0 :	Ser	Arg	Val	Tyr 590		Phe
Phe	Ser	Se 59	r As 5	p.	Tyr	Val	Lys	Lys 600	Va:	l As	n I	Lys	Aļa	Thr 605	Glu	Ala	Ala
Met	Phe 610	Le	u Ģļ	у 7	rp	Val	Glu 615	Gln	Let	ı, Va	1 1		Asp 620	Phe	Thr	Asp	Glu
Thr 625	Ser	Gl	ı Va	1 5	Ser	Thr 630	Thr	Asp	-Lys	Il		Ala 535	Asp	Ile	Thr	Ile	Ile 640
Ile	Pro	Ту	r Il	e (	31y 545	Pro	Ala	Leu	Asn	- Ile 656	e -G	Sły	Asn	Met	Leu	Tyr 655	Lys
Asp	Asp	Phe	66	1 G	Sly	Ala	Leu	Ile	Phe 665	Se	r G	Sly .	Ala	Val	Ile 670	Leu	Leu
Glu	Phe	11e	Pro	G	lu	Ile	Ala	Ile 680	Pro	Val	L	eu (	Gly	Thr 685	Phe	Ala	Leu
Val	Ser 690	Tyr	Ile	e A	la.	Asn	Lys 695	Val	Leu	Thi	· v		Gln 700	Thr	Ile	Asp	Asn
Ala 705	Leu	Ser	Lys	A	rg į	Asn 710	Glu	Lys	Trp	Asp		lu ' 15	Val	Tyr	Lys	Tyr	Ile 720
/al	Thr	Asn	Trp	7.	eu 2 25	Ala	Lys	Val	Asn	Thr 730		ln :	Ile	Asp	Leu	Ile 735	Arg
Lys	Lys	Met	Lys 740	G	lu A	Ala	Leu	Glu	Asn 745	Gln	A	la (	Glu .	Ala	Thr 750	Lÿs	Ala
le	Ile	Asn 755	Tyr	G.	ln 1	ſyr .	Asn	Gln 760	Tyr	Thr	G	lu (		Glu 765	Lys	Asn	Asn
le	Asn 770	Phe	Asn	1	le A	Asp	Asp 775	Leu	Ser	Ser	L		Leu /	Asn	Glu	Ser	Ile
sn 185	Lys	Ala	Met	11	le A	sn 90	Ile	Asn	Lys	Phe		eu <i>I</i> 95	Asn (	Gln	Cys	Ser	Val 800
er	Tyr	Leu	Met	A 8 0	sn S 05	er 1	Met	Ile	Pro	Tyr 810	G.	ly v	/al 1	Lys	Arg	Leu 815	Glu

Asp	Phe	e Asp	Ala 820	Ser	Leu	Lys	s As	P Ala 825		ı Lei	ı Lys	ту:	r Ile 830		r Asp	•
Asn	Arg	Gly 835	Thr	Leu	Ile	Gly	/ Gl:	n Val	l Asp	) Arc	J Leu	Lys 845		Ly:	s Val	
Asn	Asn 850	Thr	Leu	Ser	Thr	Asp 855	Ile	e Pro	Phe	Gln	Leu 860		Lys	Ty:	r Val	
Asp 865	Asn	Gln	Arg	Leu	Leu 870	Ser	Thi	Phe	Thr	Glu 875		Ile	Lys	Sei	Arg 880	
Pro	Gly	Pro	Glu	Thr 885	Leu	Cys	Gly	⁄ Ala	Glu 890		Val	Asp	Ala	Let 895	ı Gln	
Phe	Val	Cys	Gly 900	Asp	Arg	Gly	Phe	Tyr 905		Asn	Lys	Pro	Thr 910	_	Tyr	
Gly	Ser	Ser 915	Ser	Arg	Arg	Ala	Pro 920		Thr	Gly	Ile	Val 925		Glu	Cys	÷
Cys	Phe 930	Arg	Ser	Cys	Asp	Leu 935	Arg	Arg	Leu	Glu	Met 940		Cys	Ala	Pro	
Leu 945	Lys	Pro	Ala	Lys	Ser 950	Ala	Glu	Ala	•	•						
(2)	INFO	RMAT	CION	FOR	SEQ	ID 1	: 07	15:								
	(ii)	(A (B (C (D MOL FEA (A	QUENC ) LE ) TY ) ST ) TO ECUL TURE ) NA ) LO	NGTH PE: RANE POLO E TY : ME/K	I: 27 nucl EDNE GY: PE:	24 leic ESS: line DNA	ase acidoul ar .(ger	pai: d ble						,		
	(xi)	SEQ	UENCI	E DE	SCRI	PTIO	N: S	EQ I	D NC	): 15	i :					
ATG Met	CAG ' Gln '	TTC ( Phe '	GTG /	AAC A Asn 1	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	4
GTT ( /al .	GAC Asp	ATT ( Ile )	GCC T Ala T 20	rac / ryr :	ATC I	AAA Lys	ATT Ile	CCA Pro 25	AAC Asn	GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	9
TG /	AAG ( Lys )	GCT T Ala F 35	TTC A	AG A	ATT ( [le	CAT . His .	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
AT A	ACA 1 Thr E 50	TTT A	CG A	AC C	CCG (	GAA ( Glu ( 55	GAA Glu	GGA Gly	GAC Asp	TTG . Leu .	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	19:
CA A la I 65	ys C	CAG G Sln V	TG C al P	CA G	70 TT 7	CA Ser	TAC Tyr	TAC (	GAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80	240
AC A	AC G	AG A	AG G	AT A	AC I	AC C	CTG	AAG (	GGA (	GTG A	ACC A	AAA	TTA	TTC	GAG	286

CG Ar	T AT	TA Ty:	T TC r Se 10	r in	T GA r As	C CTO P Let	GGG GGI	C CG: / Arc	g Met	CTO	CTC Lei	ACC Thr	TCA Ser 110	: Ile	GTC Val	336
CG	C GG/ g Gly	A ATO / Ile 11:	Pro	A TT	T TGO	G GGT O Gly	GG( Gl <sub>y</sub> 120	' Sei	r ACC	ATI	GAC Asp	ACG Thr 125	Glu	TTG Leu	AAG Lys	384
GT: Val	T AT1	: Asp	ACT Thi	r AAC r Asi	TGC Cys	C ATT Ile 135	Asn	GT(	ATC	CAA Gln	CCA Pro 140	Asp	GGT Gly	AGC Ser	TAC	.432
AGA Arc	, Ser	GAA Glu	GA/	A CTI	AAC Asr 150	CTC Leu	GTA Val	ATC Ile	ATC	GGG Gly 155	CCC	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TT1 Phe	GAC Glu	TGC Cys 165	Lys	AGC Ser	TTT Phe	GGC	CAC His	Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	. GIA	TCT Ser	ACT	CAG Gln	TAC Tyr 185	Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	A'AG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCÁ Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	. 864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA . Lys !	Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT . Phe	AAA ( Lys !	GTA   Val   360	CTT Leu	AAC Asn	AGA Arg	Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104

TT Ph	T GA e As 37	p Ly	AA GO 's Al	CC GT .a Va	A TT 1 Ph	T AAG e Lys 375	s Ile	A AA:	T ATA	A GTA	CCT Pro	Lys	G GTA	A AA' L As:	T TAC	1152	
AC. Th: 38	r Il	A TA	T GA	T GG	A TT y Pho 390	e Asr	TTA Leu	AGA Arg	TAA A	ACA Thr 395	Asr	TTA Leu	A GCA	A GC	A AAC a Asn 400	1200	
TT'	T AA' e Asi	r GG n Gl	T CA y Gl	A AA n As 40	n Thi	A GAZ r Glu	ATT	AAT Asr	AAT Asn 410	Met	AAT Aşn	TTI Phe	C ACT	Lys 415	A CTA S Leu	, 1248	
AAJ Lys	A AA: s Asi	r TT 1 Ph	T AC e Th 42	r Gl	A TTO	TTI Phe	GAA Glu	TTT Phe 425	Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	Va]	A AGA Arg	1296	
GG( Gly	ATA / Ile	AT. 11: 43:	e Th	T TC: r Sei	r AAA r Lys	ACT Thr	Lys 440	Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	Tyr	AA] Asr	AAG Lys	1344	
ATC	GAA Glu 450	Gl	r CG	T TGO	GAT Asp	GGG Gly 455	Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	Ile	AAA Lys	GTT Val	1392	
AAT Asn 465	Asn	Tr	G GAO	C TTC D Leu	TTT Phe 470	Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440	
GAT Asp	CTA Leu	AAT Asr	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488	
GCA Ala	GAA Glu	GA# Glu	AAT Asn 500	lle	AGT Ser	TTA Leu	GAT Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536	
TTT Phe	AAT Asn	TIT Phe 515	Asp	AAT Asn	GAA Glu	CCT Pro	GAA Glu 520	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	1584	
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632	
CCT Pro 545	AAT Asn	GGA Gly	AAA Lys	AAG Lys	TAT Tyr 550	GAG Glu	TTA Leu	GAT Asp	AAA Lys	TAT Tyr 555	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 560	1680	
CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 565	TTT Phe	GAA Glu	CAT His	GGT. Gly	AAA Lys 570	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr	1728	
TAA neA	TCT Ser	GTT Val	AAC Asn 580	GAA Glu	GCA Ala	TTA Leu	Leu .	AAT Asn 585	CCT Pro	AGT Ser	CGT Arg	GTT Val	TAT Tyr 590	ACA Thr	TTT Phe	1776	
TTT Phe	TCT Ser	TCA Ser 595	GAC Asp	TAT Tyr	GTA Val	Lys	AAA Lys 600	GTT Val	AAT Asn	AAA Lys	GCT Ala	ACG Thr 605	GAG Glu	GCA Ala	GCT Ala	1824	
Met	TTT Phe 610	TTA Leu	GGC Gly	TGG Trp	GTA Val	GAA Glu 615	CAA ( Gln )	TTA Leu	GTA Val	Tyr	GAT Asp 620	TTT Phe	ACC Thr	GAT Asp	GAA Glu	1872	
ACT Thr 625	AGC Ser	GAA Glu	GTA Val	AGT Ser	ACT . Thr '	ACG (	GAT A	AAA Lys	Ile .	GCG Ala 635	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 640	1920	

ATT Ile	CCA Pro	TA:	r AT.	A GG e Gl 64	y Pro	r GCT o Ala	TTI Let	A AAT 1 Ast	ATA 1 Ile 650	Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 655	-		1968
GAT Asp	GAT Asp	TTT Phe	GT Va: 660	l Gly	r GCT / Ala	r TTA a Leu	ATA Ile	TTI Phe 665	Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 670	CTG Leu	TTA		2016
GAA Glu	TTT Phe	ATA Ile 675	CCI Pro	A GAC	ATŢ	GCA Ala	ATA Ile	Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala	CTT		2064
GTA Val	TCA Ser 690	Tyr	ATI	GCG Ala	TAA 3	AAG Lys 695	Val	CTA Leu	ACC	GTT Val	CAA Gln 700	Thr	ATA Ile	GAȚ Asp	AAT Asn		2112
GCT Ala 705	TTA Leu	AGT Ser	Lys	AGA Arg	AAT Asn 710	'Glu	AAA Lys	TGG	GAT Asp	GAG Glu 715	GTC Val	TAT Tyr	AAA Lys	TAT Tyr	ATA Ile 720	,	2160
GTA Val	ACA Thr	AAT Asn	TGG	TTA Leu 725	Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT	GAT Asp	CTA Leu	ATA Ile 735	AGA Argʻ		2208
AAA Lys	AAA Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala		2256
ATA Ile	ATA Ile	AAC Asn 755	TAT	CAG Gln	TAT	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AAT Asn		2304
ATT Ile	AAT Asn 770	TTT Phe	AAT Asn	ATT Ile	GAT Asp	GAT Asp 775	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	TAA Asn	GAG Glu	TCT	ATA Ile		2352
			ATG Met											Ser			2400
TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GAA Glu		2448
			GCT Ala 820													•	2496
			ACT Thr			Gly											2544
			CTT Leu		Thr												2592
			AGA Arg	Leu					Thr								2640
			AAA Lys					Ile									2688
ATT Ile	GAT Asp	ACA Thr	CAT . His . 900	AAT / Asn /	AGA Arg	ATT I	Lys	GAT Asp 905	GAA Glu	TTA Leu	TGA *						2724

#### (2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 908 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
45
45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr . 130 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ille Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

305	Ini	inz	MIG	261	310	GIN	Tyr	met	Lys	Asn 315	Val	Phe	Lys	Glu	Lys 320
_	_		_		_										

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390, 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445

Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460

Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480

Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495

Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510

Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525

Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540

Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555

Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575

Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590

Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620

Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640

Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650 655 WO 98/07864 PCT/GB97/02273

- 79 -

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
705 710 715 720

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
725 730 735

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Lys Asn Asn 755 760 765

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
770 780

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg 865 870 875 880

Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp 885 890 895

Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu

- (2) INFORMATION FOR SEQ ID NO: 17:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 3042 base pairs(B) TYPE: nucleic acid
    - (C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
  - (ix) FEATURE:
    - (A) NAME/KEY: CDS
    - (B) LOCATION:1..3042
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1

GTI Val	GAC Asp	ATT Ile	GCC Ala 20	Туг	ATC Ile	AAA Lys	ATI	CCA Pro 25	AAC Asn	GCC Ala	GGC	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
GTG Val	AAG Lys	GCT Ala 35	Phe	Lys	ATT	CAT His	AAC Asn 40	Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
GAT Asp	ACA Thr 50	Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
GCA Ala 65	Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	Ser	TAC Tyr	TAC Tyr	ĠAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC, Ser	ACA Thr 80	240
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC, Tyr	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu	. 288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
					TCT Ser											576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
					GCA Ala						_					672
					CAT His 230											720
					AAC Asn											768
					GAA Glu											816
TTT Ph					CAG Gln											864

AAC Lys	777 Phe 290	e Ly	A GA	TA TA	TT GC .e Al	A AG a Se 29	r Th	A CT Ir Le	G AA u As	C AA n Ly	G GC s Al 30	a Ly	G TC s Se	C AI r Il	T GTG e Val		912
GGT Gly 305	Thr	AC Th	T GC r Al	T TC	A TI r Le 31	u Gl	G TA n Ty	T AT	G AA t Ly	A AA' s As: 31:	n Va	T TT	T AA e Ly	A GA s Gl	G AAA u Lys 320		960
TAT Tyr	CTC Leu	Le	A TC u Se	T GA r Gl 32	u As	T AC.	A TC r Se	T GG	A AA y Ly: 33	s Phe	T TC	G GT r Va	A GA' l As <sub>l</sub>	T AA p Ly 33	A TTA s Leu 5		1008
AAA Lys	TTT Phe	GA' Asj	T AA p Ly 34	s Le	A TA u Ty	C AA	A ATO	G TT Let 34!	ı Thi	A GAO	AT'	T TAC	C AC	r Gl	G GAT u Asp		1056
AAT Asn	TTT Phe	GT: Val 355	l Ly:	G TT s Pho	T TT	r aaj e Lys	360	l Let	AA 1 12A L	C AGA	A AAJ	A ACA 5 Thi 365	Ty	TTO Let	G AAT Asn		1104
Phe	370	Lys	s Ala	a Val	l Phe	375	Ile	e Asr	ı Ile	≥ Val	380	) Lys	: Val	l Ası	TAC Tyr		1152
Thr 385	Ile	Туг	Asp	Gly	Phe 390	Asn	Leu	Arg	Asn	395	Asr	. Leu	Ala	Ala	AAC ASD 400		1200
Phe	Asn	Gly	Gln	405	Thr	· Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415		ı	1248
Lys	Asn	Phe	Thr 420	Gly	Leu	Phe	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	AGA Arg		1296
GGG GG	Ile	11e 435	Thr	Ser	Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys		1344
	Glu 450	Gly	Arg	Cys	Asp	Gly 455	Ala	Leu	Asn	Asp	Leu 460	Сув	Ile	Lys	Val	,	1392
AAT A Asn A 465	Asn '	Trp	Asp	Leu	Phe 470	Phe	Ser	Pro	Ser	Glu 475	Asp	Asn	Phe	Thr	Asn 480		1440
GAT (	Leu /	Asn	Lys	Gly 485	Glu	Glu	Ile	Thr	Ser 490	Asp	Thr	Asn	Ile	Glu 495	Ala		1488
GCA C	3lu (	Glu	Asn 500	Ile	Ser	Leu	Asp	Leu 505	Ile	Gln	Gln	Tyr	Tyr 510	Leu	Thr		1536
TTT A	sn I	Phe	GAT Asp	AAT Asn	GAA Glu	Pro	GAA Glu 520	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser		1584
	sp I	lle .	Ile	Gly	Gln	Leu 535	Glu	Leu	Met	Pro	Asn 540	Ile	Glu	Arg	Phe		1632
CCT A Pro A 545	AT G sn G	GA A	AAA . Lys	Lys	TAT Tyr 550	GAG Glu	TTA Leu	GAT . Asp	Lys	TAT Tyr 555	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 560		1680

CTI Leu	CGT Arg	GCT	CAA Gln	GAA Glu 565	ı Phe	GAA Glu	CAT	GGT Gly	AAA Lys 570	Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr		1728
AAT Asn	TCT Ser	GTI Val	AAC Asn 580	GAA Glu	GCA Ala	TTA Leu	TTA Leu	AAT Asn 585	Pro	AGT	CGT Arg	CTT Val	TAT Tyr 590	ACA Thr	TTT Phe		1776
TTT Phe	TCT	TCA Ser 595	GAC Asp	TAT	GTA Val	'AAG Lys	AAA Lys 600	Val	AAT Asn	AAA Lys	GCT Ala	ACG Thr 605	GAG Glu	GCA Ala	GCT Ala		1824
ATG Met	TTT Phe 610	TTA Leu	GGC	TGG	GTA Val	GAA Glu 615	CAA Gln	TTA Leu	GTA Val	TAT	GAT Asp 620	TTT Phe	ACC Thr	GAT Asp	'GAA Glu		1872
ACT Thr 625	AGC Ser	GAA Glu	GTA Val	AGT Ser	ACT Thr 630	ACG Thr	GAT Asp	AAA Lys	ATT Ile	GCG Ala 635	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 640		1920
ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 645	CCT Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 650	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 655	AAA Lys	ı	1968
GAT Asp	GAT Asp	TTT Phe	GTA Val 660	GGT Gly	GCT Ala	TTA Leų	ATA Ile	TTT Phe 665	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 670	CTG Leu	TTA Leu		2016
GAA Glu	TTT Phe	ATA Ile 675	CCA Pro	GAG Glu	ATT Ile	GCA Ala	ATA Ile 680	CCT	GTA Val	TTA Leu	GGT Gly	ACT Thr 685	TTT Phe	GCA, Ala	CTT Leu		2064
			ATT Ile														2112
			AAA Lys														2160
GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg		2208
			AAA Lys 740													•	2256
			TAT Tyr														2304
			AAT Asn														2352
			ATG Met														2400
			ATG Met														2448
GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT Tyr	GAT Asp		2496

- 83 -

AAT Asr	r AGA	GGA Gly 835	Thr	TTA Lev	ATI	GGT Gly	CAA Gln 840	Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val		2544
AA1 Asr	TAAT Asn 850	Thr	CTI Leu	AGI Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val		2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCA Ser	GGC Gly 880		2640
CTG Leu	AAT Asn	TCC Ser	CCG Pro	GGT Gly 885	Ala	GCT Ala	CAT His	TAT Tyr	GCG Ala 890	CAA Gln	CAC His	GAT Asp	GAA Glu	GCC Ala 895	GTA Val		2688
GAC Asp	AAC Asn	AAA Lys	TTC Phe 900	AAC Asn	AAA Lys	GAA Glu	CAA Gln	CAA Gln 905	AAC Asn	GCG Ala	TTC Phe	TAT Tyr	GAG Glu 910	ATC Ile	TTA Leu		2736
CAT	TTA Leu	CCT Pro 915	AAC Asn	TTA Leu	AAC Asn	GAA Glu	GAA Glu 920	CAA Gln	CGA Arg	AAC Asn	GCC Ala	TTC Phe 925	ATC Ile	CAA Gln	AGT Ser		2784
TTA Leu	AAA Lys 930	GAT Asp	GAC Asp	CCA Pro	AGC Ser	CAA Gln 935	AGC Ser	GCT Ala	AAC Asn	CTT Leu	TTA Leu 940	GCA Ala	GAA Glu	GCT Ala	AAA Lys		2832
AAG Lys 945	CTA Leu	AAT Asn	GAT Asp	GCT Ala	CAG Gln 950	GCG Ala	CCG Pro	AAA Lys	GTA Val	GAC Asp 955	AAC Asn	AAA Lys	TTC Phe	AAC Asn	AAA Lys 960	,	2880
GAA Glu	CAA Gln	CAA Gln	AAC Asn	GCG Ala 965	TTC Phe	TAT Tyr	GAG Glu	ATC Ile	TTA Leu 970	CAT His	TTA Leu	CCT Pro	AAC Asn	TTA Leu 975	AAC Asn		2928
GAA Glu	GAA Glu	CAA Gln	CGA Arg 980	AAC Asn	GCC Ala	TTC Phe	ATC Ile	CAA Gln 985	AGT Ser	TTA Leu	AAA Lys	GAT Asp	GAC Asp 990	CCA Pro	AGC Ser		2976
	AGC Ser					Ala		Ala					Asp				3024
	CCG Pro 1010	Lys	_		TAG												3042

## (2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1014 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp	Thr 50		Thr	. Asr	Pro	Glu 55	Glu	Gly	Asp	Leu	Asn 60	Pro	Pro	Pro	Gl
Ala 65	Lys	Gln	Val	Pro	Val 70		Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	Th:
Asp	Asn	Glu	Lys	Asp 85	Asn	Tyr	Leu	Lys	Gly 90		Thr	Lys	Leu	Phe 95	Gl
Arg	Ile	Туг	Ser 100		Asp	Leu	Gly	Arg 105	Met	Leu	Leu	Thr	Ser 110	Ile	Va:
Arg	Gly	Ile 115	Pro	Phe	Trp	Gly	Gly 120	Ser	Thr	Ile	Asp	Thr 125	Glu	Leu	Lys
Val	Ile 130	Asp	Thr	Asn	Cys	Ile 135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser	Ту
Arg 145	Ser	Glu	Glu	Leu	Asn 150	Leu	Val	Ile	Ile	Gly 155	Pro	Ser	Ala	Asp	11e
Ile	Gln	Phe	Glu	Cys 165	Lys	Ser	Phe	Gly	His 170	Glu	Val	Leu	Asn	Leu 175	Thi
Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln ·	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe
Thr	Phe	Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thr	Asn 205	Pro	Leu	Let
Gly	Ala 210	Gly	Lys	Phe	Ala	Thr 215	Asp	Pro	Ala	Val	Thr 220	Leu	Ala	His	Glu
Leu- 225	-I-le-	His	-Ala-	-Gly	His- 230	Arg	-Leu~	Tyr-	-Gly	Tle- 235	Ala <sup>-</sup>	I'l'e-	-Asn	Pro	Asr 240
Arg	Val	Phe	Lys	Val 245	Asn	Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	Leu
Glu	Val	Ser	Phe 260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys
Phe	Ile	Asp 275	Ser	Leu	Gln	Glu	Asn 280	Glu	Phe	Arg	Leu	Tyr 285	Tyr	Tyr	Ası
Lys	Phe 290	Lys	Asp	Ile	Ala	Ser 295	Thr	Leu	Asn	Lys	Ala 300	Lys	Ser	Ile	Va]
Gly 305	Thr	Thr	Ala	Ser	Leu 310	Gln	Tyr	Met	Lys	Asn 315	Val	Phe	Lys	Glu	Lys 320
Tyr	Leu	Leu	Ser	Glu 325	Asp	Thr	Ser	Gly	Lys 330	Phe	Ser	Val	Asp	Lys 335	Leu
Lys	Phe	Asp	Lys 340	Leu	Tyr	Lys	Met	Leu 345	Thr	Glu	Ile	Tyr	Thr 350	Glu	Asţ
Asn	Phe	Val 355	Lys	Phe	Phe	Lys	Val 360	Leu	Asn	Arg	Lys	Thr 365	Tyr	Leu	Asr
Phe	Asp 370	Lys	Ala	Val	Phe	Lys 375	Ile	Asn	Ile	Val	Pro 380	Lys	Val	Asn	Туз
Thr 385	Ile	Tyr	Asp	Gly	Phe 390	Asn	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	Asr 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 490 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu · 615 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 650 Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 680 675 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn
755 760 765

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 780

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu . 815

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly 865 870 875 880

Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val 885 890 895

Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu 900 905 910

His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 915 920 925

Leu Lys-Asp Asp Pro Ser-Gln-Ser-Ala Asn-Leu-Leu-Ala Glu Ala Lys 930 935

Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 945 955 960

Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn 965 970 975

Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 980 985 990

Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000 1005

Ala Pro Lys Val Asp \* 1010

#### (2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3509 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION:1..3509
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

- 87 -

ATO Met	Pro	A GT	T AC 1 Th	r Il	A AA e As 5	T AA' n Ası	T TT:	T AA:	T TAT	r Ası	T GA: n Ası	r cc:	r ATT	GA:	TAAT Asn	÷	48
AAT Asn	AAT Asr	TAT	e Il	T AT e Me 0	G ATO	G GAO	G CCT	CCA Pro 25	Phe	r GC0 ≥ Ala	G AGA	A GGT g Gly	ACC Thi	Gly	G AGA / Arg		96
TAT	TAT	Lys	s Al	T TT	T AA e Ly:	A ATO	ACA Thr	Asp	CGT Arg	T ATT	TGC Trp	ATA Ile 45	: Ile	CCC Pro	G GAA		144
AGA Arg	TAT Tyr 50	Thi	r TT	T GG e Gly	а тал у Тул	Lys 55	Pro	GAG	GAT ASP	TTI Phe	TAA T Raa S Oo	Lys	AGT Ser	TCC Ser	GGT		192
ATT Ile 65	Phe	AA] Asr	r AG	A GAT	GTT Val	. Cys	GAA Glu	TAT	TAT	GAT Asp 75	Pro	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 80		240
ACT Thr	AAT Asn	GAT Asp	Lys	A AAC S Lys 85	Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	Thr	ATG Met	ATC Ile	AAG Lys	TTA Leu 95			288
AAT Asn	AGA Arg	ATC	Lys 100	Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT Ile		336
			Ile	CCT Pro				Asp								•	384
Phe	Asn 130	Thr	Asn	ATT Ile	Ala	Ser 135	Val	Thr	Val	Asn	Lys 140	Leu	Ile	Ser	Asn		432
CCA Pro 145	GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160		480
Phe	Gly	Pro	Gly	CCA Pro 165	Val	Leu	Asn	Glu	Asn 170	Glu	Thr	Ile	Asp	Ile 175	Gly		528
lle	Gln	Asn	His 180	TTT	Ala	Ser	Arg	Glu 185	Gly	Phe	Gly	Gly	Ile 190	Met	Gln		576
1et	Lys	Phe 195	Cys	CCA Pro	Glu	Tyr	Val 200	Ser	Val	Phe	Asn	Asn 205	Val	Gln	Glu	•	624
sn	Lys 210	Gly	Ala	AGT Ser	Ile	Phe 215	Asn .	Arg	Arg	Gly	Tyr 220	Phe	Ser	Asp	Pro		672
				ATG Met													720
				GAT Asp 245				Ile									768
		Gln		ACA Thr			Ile					Leu					816

		CAA Gln 275														864
		AAA Lys														912
		TTA Leu				Ser										960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
		GGT Gly 355														1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
		TTA Leu							Ile							1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met~ 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gl'n	AAT Asn	AAA 'Lys'	GCT Ala 415	ATA Tle	1248
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	TAA RSN	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1632

AC. Th: 54	r Ph	T CO	CT C	TA G	AT AT sp Il 55	le Ar	GA GA G As	T AT p Il	A AG' e Se:	T TT r Lev 559	ı Thi	A TC: r Se:	T TC	A TT	T GAT e Asp 560		1680
GA' Asj	F GC.	A TI a Le	TA T	TA TT eu Pi 56	ne Se	T AA r As	C AA n Ly	A GT s Va	T TA: 1 Ty: 570	r Sei	A TTT	TTT	TC:	T ATO	G GAT E Asp		1728
TA:	r AT	r AA e Ly	A AC 's Ti 58	ır Al	CT AA .a As	T AA n Ly	A GTO	G GT l Va 58	l Glu	A GCA 1 Ala	GGF Gly	A TTA Leu	TT: Phe 590	a Ala	A GGT A Gly		1776
TG(	GT(	AA Ly 59	s Gl	G AT	A GT e Va	A AA l As	T GA n Ası 600	Phe	Γ GTA ⊇ Val	A ATC	GAA Glu	GCT Ala 605	Asr	AAA Lys	A AGC S Ser		1824
AAT Asn	ACT Thr 610	: Me	G GA t As	T AA p Ly	A AT	T GC e Ala 61	a Asp	T ATA	A TCI	CTA Leu	ATT Ile 620	Val	CC1	TAT Tyr	ATA : Ile		1872
Gly 625	Leu	Ala	a Le	u As	n Va. 63	l Gly	/ Asr	ı Glu	Thr	Ala 635	Lys	Gly	Asn	Phe	GAA Glu 640		1920
Asn	Ala	Phe	e Gl	u Il 64	e Ala 5	a Gly	/ Ala	Ser	1le 650	Leu	Leu	Glu	Phe	Ile 655			1968
Glu	Leu	Leu	1 Ile 66	e Pro	o Val	l Val	. Gly	Ala 665	Phe	TTA Leu	Leu	Glu	Ser 670	Tyr	Ile	•	2016
Asp	Asn	Lys 675	Ası	ı Lys	s Ile	: Ile	680	Thr	Ile	Asp	Asn	Ala 685	Leu	Thr	_		2064
Arg	Asn 690	Glu	Lys	Tr	Ser	Asp 695	Met	Tyr	Gly	TTA Leu	Ile 700	Val	Ala	Gln	Trp		2112
Leu 705	Ser	Thr	Val	. Asn	710	Gln	Phe	Tyr	Thr	ATA Ile 715	Lys	Glu	Gly	Met	Tyr 720		2160
Lys	Ala	Leu	Asn	725	Gln	Ala	Gln	Ala	Leu 730	GAA Glu	Glu	Ile	Ile	Lys 735	Tyr		2208
Arg	Tyr	Asn	Ile 740	Tyr	Ser	Glu	Lys	Glu 745	Lys		Asn	Ile	Asn 750	Ile	Asp		2256
Phe	Asn	Asp 755	Ile	Asn	Ser	Lys	Leu 760	Asn	Glu	GGT Gly	Ile	Asn 765	Gln	Ala	Ile		2304
Asp	Asn 770	Ile	Asn	Asn	Phe	11e 775	Asn	Gly	Сув		Val 780	Ser	Tyr	Leu	Met		2352
Lys 7 <b>8</b> 5	Lys	Met	Ile	Pro	Leu 790	Ala	Val	Glu	Lys	TTA Leu 795	Leu	Asp	Phe	Asp	Asn 800		2400
										GAT Asp							2448

TT(	ATT	GGA Gly	AGT Ser 820	Ala	A GAA a Glu	TAT Tyr	GAZ Glu	AAA Lys 825	Ser	AAA Lys	GTA Val	AAT Asn	AAA Lys 830	Tyr	TTG Leu		2496
AA# Lys	ACC Thr	ATT Ile 835	Met	Pro	TTT Phe	GAT Asp	CTI Leu 840	Ser	ATA : Ile	TAT	ACC	AAT Asn 845	Asp	ACA Thr	ATA Ile		2544
CTA Leu	ATA Ile 850	Glu	ATG Met	TTT Phe	AAT Asn	Lys 855	Tyr	AAT Asn	AGC Ser	GAA Glu	ATT Ile 860	Leu	AAT Asn	AAT Asn	ATT		2592
ATC Ile 865	Leu	AAT Asn	TTA Leu	AGA Arg	TAT Tyr 870	Lys	GAT Asp	AAT Asn	AAT Asn	TTA Leu 875	ATA Ile	GAT Asp	TTA Leu	TCA Ser	GGA Gly 880		2640
TAT	GGG Gly	GCA Ala	AAG Lys	GTA Val 885	Glu	GTA Val	TAT	GAT Asp	GGA Gly 890	GTC Val	GAG Glu	CTT Leu	AAT Asn	GAT Asp 895	AAA Lys		2688
AAT Asn	CAA Gln	TTT Phe	AAA Lys 900	TTA Leu	ACT Thr	AGT Ser	TCA Ser	GCA Ala 905	AAT Asn	AGT Ser	AAG Lys	ATT Ile	AGA Arg 910	GTG Val	ACT Thr		2736
CAA Gln	AAT Asn	CAG Gln 915	AAT Asn	ATC Ile	ATA Ile	TTT Phe	AAT Asn 920	AGT Ser	GTG Val	TTC Phe	CTT Lėu	GAT Asp 925	TTT Phe	AGC Ser	GTT Val		2784
AGC Ser	TTT Phe 930	TGG Trp	ATA Ile	AGA Arg	ATA Ile	CCT Pro 935	AAA Lys	TAT Tyr	AAG Lys	AAT Asn	GAT Asp 940	GGT Gly	ATA Ile	CAA Gln	AAT Asn		2832
TAT Tyr 945	ATT Tle	CAT His	AAT Asn	GAA Glu	TAT Tyr 950	ACA Thr	ATA Ile	ATT Ile	AAT Asn	TGT Cys 955	ATG Met	AAA Lys	AAT Asn	AAT Asn	TCG Ser 960		2880
GGC Gly	TGG Trp	AAA Lys	ATA Ile	TCT Ser 965	ATT Ile	AGG Arg	GGT Gly	AAT Asn	AGG Arg 970	ATA Ile	ATA Ile	TGG Trp	ACT Thr	TTA Leu 975	ATT Ile	٠	2 <b>92</b> 8
GAT Asp	ATA Ile	AAT Asn	GGA Gly 980	AAA Lys	ACC Thr	AAA Lys	TCG Ser	GTA Val 985	TTT Phe	TTT Phe	GAA Glu	TAT Tyr	AAC Asn 990	ATA Ile	AGA Arg		2976
		ATA Ile 995						Arg					Thr				3024
AAT Asn	AAT Asn 1010	TTG Leu	AAT Asn	AAC Asn	GCT Ala	AAA Lys 1015	Ile	TAT Tyr	ATT Ile	AAT Asn	GGT Gly 1020	Lys	CTA Leu	GAA Glu	TCA Ser		3072
	Thr	GAT Asp		Lys		Ile					Ala						3120
		AAA Lys	Leu		Gly					Thr					Met		3168
		TTC Phe		Ile			Thr		Leu					Ile			3216
		TAT Tyr 1075	Lys			Ser		Ser					Asp				3264

- 91 -

		CCT Pro					Lys					Phe					3312
	Lys	AAT Asn				Lys					Ser						3360
		ACA Thr			Lys					Ser					Tyr	,	3408
AGA Arg	GAT Asp	TTA Leu	TAT Tyr 1140	Ile	GGA Gly	GAA Glu	AAA Lys	TTT Phe 1145	Ile	ATA Ile	AGA Arg	AGA Arg	AAG Lys 1150	Ser	AAT Asn	:	3456
		TCT Ser 1155	Ile			Asp		Val					Tyr			,	3504
CTA Leu	GA														ı	1	3509

#### (2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1169 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly, 50

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 265 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 315 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 360 Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 375 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 390 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 410 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 505 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys

Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 535 Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 550 · 555 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 600 Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 650 Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 695 Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720 Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 840 Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly 870

Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 885 890 895

Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr 900 905 910

Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 915 920 925

Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 930 935 940

Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 945 950 955 960

Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile 965 970 975

Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 980 985 990

Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 995 1000 1005

Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1010 1015 1020

Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 1040

Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met
1045 1050 1055

Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1060 1065 1070

Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080 1085

Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly
1090 1095 1100

Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu 1105 1110 1115 1120

Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr
1125 1130 1135

Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn 1140 1145 1150

Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr 1155 1160 1165

Leu

- (2) INFORMATION FOR SEQ ID NO: 21:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2574 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)

- 95 -

#### (ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2574

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

ATC Met	Pro	GT Va	r ACA l Thi	A ATA	A AAT e Asr	r AAT 1 Asr	TTT n Phe	TAA 1 Asi	TAT Tyr	Asr	GAT Asp	CCI Pro	AT1	GAT Asp	AAT Asn	4
AA] Asr	TAA 1 Asi	AT:	T ATT	: Met	ATC Met	GAG Glu	CCI Pro	CCA Pro 25	Phe	GCG Ala	AGA Arg	GGT Gly	ACG Thr	Gly	AGA Arg	90
TAI Tyr	TAT	Lys 35	. Ala	TTI Phe	AAA Lys	ATC	ACA Thr 40	Asp	CGT Arg	ATI Ile	TGG Trp	ATA Ile 45	Ile	CCG Pro	GAA Glu	144
AGA Arg	TAT Tyr 50	Thr	TTT Phe	GGA Gly	TAT	AAA Lys 55	Pro	GAG Glu	GAT Asp	TTT Phe	AAT Asn 60	Lys	AGT Ser	TCC Ser	GGT Gly	192
ATT Ile 65	Phe	AAT Asn	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT	TAT	GAT Asp 75	Pro	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 80	240
ACT Thr	AAT Asn	GAT Asp	AAA Lys	AAG Lys 85	AAT Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC	AAG Lys	TTA Leu 95	TTT Phe	288
AAT Asn	AGA Arg	ATC Ile	AAA Lys 100	TCA Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT	336
ATA Ile	AAT Asn	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT- Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu	384
TTT Phe	AAC Asn 130	ACA Thr	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn	432
CCA Pro 145	GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160	480
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly	528
ATA Ile	CAA Gln	AAT Asn	CAT THIS	TTT Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC Gly	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln	576
ATG Met	Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	TAA Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	624
Asn			GCA . Ala :		Ile											672
GCC Ala 225				Met :					His							720

G)	C AT y Il	T AA e Ly	A GTA	A GAT l Asp 24!	p Ası	r TT/ p Let	A CCA	A ATT	CT/ Val 250	Pro	A AAT Asr	GAA Glu	AAA Lys	Lys 255	TTT Phe	. 76	8 ;
TI Ph	T AT e Me	G CA	A TC	r Thi	A GAT	C GCT Ala	T ATA	CAC Glr 265	Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	Thr	TTT	81	6
GG Gl	A GG y Gl	A CAP 7 Gli 27	A GAT n Asp	CCC Pro	C AGO Ser	T Ile	ATA Ile 280	Thr	CCI Pro	TCT	ACG Thr	GAT Asp 285	Lys	AGT Ser	ATC	86	4
TA Ty	T GAT r Ası 290	Lys	GTT Val	TTC Leu	CAA Gln	AAT Asn 295	Phe	AGA Arg	GGG	ATA Ile	GTT Val	Asp	AGA Arg	CTT Leu	AAC Asn	. 91	2
AA Ly: 30:	s Val	TTI Lei	A GTI ı Val	TGC Cys	ATA Ile 310	Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	Asn	ATT	AAT Asn	ATA Ile	TAT Tyr 320	96	0
AAI Lys	A AAl s Asr	AAA Lys	TTT Phe	Lys 325	Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	100	8
AA) Lys	TAT	AGT Ser	Ile 340	Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	105	6
ATC Met	TTT Phe	GGT Gly 355	TTT	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	,110	4
ACT Thi	AGA Arg 370	Ala	TCT	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	115	2
AAT Asn 385	Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200	0
TCT	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1241	В
TAA neA	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296	5
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344	2
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392	2
GAT Asp 465	Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	TAA neA	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440	)
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488	3
			AAA Lys 500				Pro									1536	5

GAT Asp	TTT Phe	AAT Asn 515	Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	15	84
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	16	32
ACA Thr 545	TTT Phe	CCT Pro	CTA Leu	GAT Asp	'ATA Ile 550	'AGA Arg	GAT Asp	ATA Ile	AGT Ser	TTA Leu 555	ACA Thr	TCT Ser	TCA Ser	TTT Phe	GAT Asp 560	. 16	80
GAT Asp	GCA Ala	TTA Leu	TTA Leu	TTT Phe 565	TCT Ser	AAC Asn	AAA Lys	GTT Val	TAT Tyr 570	TCA Ser	TTT Phe	TTT Phe	TCT	ATG Met 575	GAT Asp	17	28
TAT Tyr	ATT Ile	AAA Lys	ACT Thr 580	GCT Ala	AAT Asn	AAA Lys	GTG Val	GTA Val 585	GAA Glu	GCA Ala	GGA Gly	TTA Leu	TTT Phe 590	GCA Ala	GGT Gly	. 17	76
TGG Trp	GTG Val	AAA Lys 595	CAG Gln	ATA Ile	GTA Val	AAT Asn	GAT Asp 600	TTT Phe	GTA Val	ATC Ile	GAA Glu	GCT Ala 605	AAT Asn	AAA Lys	AGC Ser	18	24
AAT Asn	ACT Thr 610	ATG Met	GAT Asp	AAA Lys	ATT Ile	GCA Ala 615	GAT Asp	ATA Ile	TCT Ser	CTA Leu	ATT Ile 620	GTT Val	CCT Pro	TAT Tyr	ATA Ile	18	72
GGA Gly 625	TTA Leu	GCT Ala	TTA Leu	AAT Asn	GTA Val 630	GGA Gly	AAT Asn	ĠAA Glu	ACA Thr	GCT Ala 635	AAA Lys	GGA Gly	AAT Asn	TTT Phe	GAA Glu 640	19	20
AAT Asn	GCT Ala	TTT Phe	GAG Glu	ATT Ile 645	GCA Ala	GGA Gly	GCC Ala	AGT Ser	ATT Ile 650	CTA Leu	CTA Leu	GAA Glu	TTT Phe	ATA Ile 655	CCA Pro	19	68
GAA Glu	CTT Leu	TTA Leu	ATA Ile 660	CCT Pro	GTA Val	GTT Val	GGA Gly	GCC Ala 665	TTT Phe	TTA Leu	TTA Leu	GAA Glu	TCA Ser 670	TAT Tyr	ATT Ile	20	16
GAC Asp	AAT Asn	AAA Lys 675	AAT Asn	AAA Lys	ATT Ile	ATT Ile	AAA Lys 680	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 685	TTA Leu	ACT Thr	AAA Lys	20	64
AGA Arg	AAT Asn 690	GAA Glu	AAA Lys	TGG Trp	AGT Ser	GAT Asp 695	ATG Met	TAC Tyr	GGA Gly	TTA Leu	ATA Ile 700	GTA Val	GCG Ala	CAA Gln	TGG , Trp	21	12
CTC Leu 705	TCA Ser	ACA Thr	GTT Val	AAT Asn	ACT Thr 710	CAA Gln	TTT Phe	TAT Tyr	ACA Thr	ATA Ile 715	AAA Lys	GAG Glu	GGA Gly	ATG Met	TAT Tyr 720	21	60
AAG Lys	GCT Ala	TTA Leu	Asn	TAT Tyr 725	CAA Gln	GCA Ala	CAA Gln	GCA Ala	TTG Leu 730	GAA Glu	GAA Glu	ATA Ile	ATA Ile	AAA Lys 735	TAC Tyr	22	8 0
AGA Arg	TAT Tyr	AAT Asn	ATA Ile 740	TAT Tyr	TCT Ser	GAA Glu	AAA Lys	GAA Glu 745	AAG Lys	TCA Ser	AAT Asn	ATT Ile	AAC Asn 750	ATC Ile	GAT Asp	22	56
TTT Phe	AAT Asn	GAT Asp 755	ATA Ile	AAT Asn	TCT Ser	AAA Lys	CTT Leu 760	AAT Asn	GAG Glu	GGT Gly	ATT Ile	AAC Asn 765	CAA Gln	GCT Ala	ATA Ile	23	04
GAT Asp	AAT Asn 770	ATA Ile	AAT Asn	AAT Asn	Phe	ATA Ile 775	AAT Asn	GGA Gly	TGT Cys	TCT Ser	GTA Val 780	TCA Ser	TAT Tyr	TTA Leu	ATG Met	23	52

AAA Lys 785	AAA Lys	ATG Met	ATT	CCA Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA Leu	AAT Asn	TAT Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448
			AGT Ser 820													2496
			ATG Met													2544
			ATG Met									1				2574

- (2) INFORMATION FOR SEQ ID NO: 22:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 858 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear

  - (ii) MOLECULE TYPE: protein
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

٨	let 1	Pro	Val	Thr	Ile 5	Asn	Asn	Phe	Asn	Tyr 10	Asn	Asp	Pro	Ile	Asp 15	Asn
A	İsn	Asn	Ile	Ile 20	Met	Met	Glu	Pro	Pro 25	Phe	Ala	Arg	Gly	Thr 30	-	Arg
I	yr	Tyr	Lys 35	Ala	Phe	Lys	Ile	Thr 40	Asp	Arg	Ile	Trp	Ile 45	Ile	Pro	Ģlu
A	rg	Tyr 50	Thr	Phe	Gly	Tyr	Lys 55	Pro	Glu	Asp	Phe	Asn 60	Lys	Ser	Ser	Gly
	1e 65	Phe	Asn	Arg	Asp	Val 70	Cys	Glu	Tyr	Tyr	Asp 75	Pro	Asp	Tyr	Leu	Asn 80
T	hr	Asn	Asp	Lys	Lys 85	Asn	Ile	Phe	Leu	Gln 90	Thr	Met	Ile	Lys	Leu 95	Phe
A	sn	Arg	Ile	Lys 100	Ser	Lys	Pro	Leu	Gly 105	Glu	Lys	Leu	Leu	Glu 110	Met	Ile
I	le	Asn	Gly 115	Ile	Pro	Tyr	Leu	Gly 120	Asp	Arg	Arg	Val	Pro 125	Leu	Glu	Glu
P	he	Asn 130	Thr	Asn	Ile	Ala	Ser 135	Val	Thr	Val	Asn	Lys 140	Leu	Ile	Ser	Asn
	ro 45	Gly	Glu	Val	Glu	Arg 150	Lys	Lys	Gly	Ile	Phe 155	Ala	Asn	Leu	Ile	Ile 160
P	he	Gly	Pro	Gly	Pro 165	Val	Leu	Asn	Glu	Asn 170	Glu	Thr	Ile	Asp	Ile 175	Gly
I	le	Gln	Asn	His 180	Phe	Ala	Ser	Arg	Glu 185	Gly	Phe	Gly	Gly	Ile 190	Met	Gln

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 200 Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 215 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 1 12 1 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 250 Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 310 315 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 360 365 Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 375 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 440 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 535 530

 545
 550
 555
 560

 Asp Ala Leu Leu Phe 565
 Ser Asn Lys Val Tyr Ser Phe Phe Ser Met 575
 Asp 575

 Tyr Ile Lys Thr 580
 Ala Asn Lys Val Val 585
 Glu Ala Gly Leu Phe 590
 Ala Gly 590

 Trp Val Lys Gln Ile Val Asn Asp Asp Phe Val Ile Glu Ala Asn Lys Ser 605
 Asn Lys Ser 600
 Fro 600
 Phe Val Ile Glu Ala Asn Lys Ser 605

 Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 620
 Fro 620
 Fro 635
 Fro 635

 Gly Leu Ala Leu Asn Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 640
 Fro 655
 Fro 655

 Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 655
 Fro 655

 Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660
 Fro 675

 Asp Asn Lys Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 680

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser L u Thr Ser Ser Phe Asp

Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720

Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp

695

Lys.Ala\_Leu\_Asn Tyr.Gln\_Ala\_Gln\_Ala\_Leu\_Glu\_Glu\_Ile\_Ile Lys Tyr 725 730 735

Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740 745 750

Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755 760 765

Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 780

Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 785 790 795

Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 810 815

Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 820 825 830

Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 835 840 845

Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser 850 855

- (2) INFORMATION FOR SEQ ID NO: 23:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 1644 base pairs
    - (B) TYPE: nucleic acid(C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..1644

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

													1				
ATG Met 1	Pro	GTT Val	ACA Thr	ATA Ile 5	Asn	AAT	TTT Phe	AAT Asn	TAT Tyr 10	AAT Asn	GAT Asp	CCT Pro	ATT	GAT Asp 15	AAT Asn	,	48
AAT Asn	AAT Asn	ATT	ATT Ile 20	Met	ATG Met	GAG Glu	CCT Pro	CCA Pro 25	TTT Phe	GCG Ala	AGA Arg	GGT Gly	ACG Thr 30	GĠG Gly	AGA Arg		96
TAT Tyr	TAT Tyr	AAA Lys 35		TTT Phe	AAA Lys	ATC Ile	ACA Thr 40	GAȚ Asp	CGT Arg	ATT Ile	TGG Trp	ATA Ile 45	ATA Ile	CCG Pro	GAA Glu	ø	144
AGA Arg	TAT Tyr 50	ACT Thr	TTT	GGA Gly	TAT Tyr	AAA Lys 55	CCT Pro	GAG Glu	GAT Asp	TTT	AAT Asn 60	AAA Lys	AGT Ser	TCC Ser	GGT Gly		192
ATT Ile 65	TTT Phe	AAT Asn	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT Tyr	TAT Tyr	GAT Asp 75	CCA Pro	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 80		240
ACT Thr	AAT Asn	GAT Asp	AAA Lys	AAG Lys 85	AAT Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC Ile	AAG Lys	TŤA Leu 95	TTT Phe		288
AAT Asn	AGA Arg	ATC Ile	AAA Lys 100	TCA Ser	Lys	CCA Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT Ile		336
ATA Ile	AAT Asr	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu		384
TTT Phe	AAC Asn 130	ACA Thr	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn		432
CCA Pro 145	GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA' Ile 160		480
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly		528
ATA Ile	CAA Gln	AAT Asn	CAT His 180	TTT Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC Gly	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln		576
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu		624
AAC Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT Arg	GGA Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro		672

GCC Ala 225	Leu	ATA Ile	TTA Leu	ATO Met	CAT His 230	Glu	CTI Leu	ATA Ile	CAT His	GTT Val 235	Leu	CAT His	GGA Gly	TTA Leu	TAT Tyr 240	720
GGC	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	Asp	TTA Leu	CCA Pro	ATT Ile	GTA Val 250	Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	Phe	768
TTT Phe	ATG Met	CAA Gln	TCT Ser 260	Thr	GAT Asp	GCT	ATA Ile	CAG Gln 265	Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	ACA Thr	TTT Phe	816
GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC	AGC Ser	ATC Ile	ATA Ile 280	Thr	CCT	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC	864
TAT Tyr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lyş	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	Glu	AGT Ser	Phe	GAT Asp	Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met-	TTT -Phe-	GGT Gly 355	TTT -Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Tle	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	TAA Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
			GCT Ala 420													1296
			ATG Met													1344
			GAA Glu		Leu											1392
			TCT Ser	Lys												1440
			AAT Asn													1488

WO 98/07864 PCT/GB97/02273

- 103 -

	AGT Ser								1536
	AAT Asn 515								1584
	TTT Phe						 	 	 . 1632
 	CCT Pro	-			·			ı	1644

- (2) INFORMATION FOR SEQ ID NO: 24:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 548 amino acids
    - (B) TYPE: amino acid
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg
20 25 30

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu
35 40 45

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly
165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
180 185 190

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205

Asn	Lys 210		Ala	Ser	: Ile	Phe 215		Arg	Arg	Glγ	Tyr 220	Phe	Ser	Asp	Pro
212			I.e.	. Met	His			Tle	Hie	Val		Wie	Gly	Leu	т
225	peu	116	рсс		230			***		235	пец	1115	GIY		, 240
Gly	Ile	Lys	Val	Asp 245	,	Leu	Pro	Ile	Val 250		Asn	Glu		Lys 255	Phe
Phe	Met	Gln	Ser 260		Asp	Ala	Ile	Gln 265		Glu	Glu	Leu	Tyr 270	Thr	Phe
Gly	Gly	Gln 275	Asp	Pro	Ser	Ile	Ile 280	Thr	Pro	Ser	Thr	Asp 285	Lys	Ser	Ile
Tyr	Asp 290	Lys	Val	Leu	Gln	Asn 295	Phe	Arg	Gly	Ile	Val 300	Asp	Arg	Leu	Asr
Lys 305	Val	Leu	Val	Cys	Ile 310	Ser	Asp	Pro	Asn	Ile 315	Asn	Ile	Asn	Ile	Tyr 320
Lys	Asn	Lys	Phe	Lys 325	Asp	Lys	Tyr	Lys	Phe 330	Val	Glu	Asp	Ser	Glu 335	Gly
Lys	Tyr	Ser	11e 340	Asp	Val	Glu	Ser	Phe 345	Asp	Lys	Leu	Tyr	Lys 350	Ser	Lev
Met	Phe	Gly 355	Phe	Thr	Glu	Thr	Asn 360	Ile,	Ala	Glu	Asn	Tyr 365	Lys	Ile	Lys
Thr	Arg 370	Ala	Ser	Туг	-Phe-	Ser- 375	Asp	Ser	Leu	-Pro	-Pro-	Val-	Lys	Ile	Lys
Asn 385	Leu"	-Leu	Asp	-Asn-	-Glu- 390	-I-l·e-	Tyr-	-Thr-	-Ile	-Glu- 395	-Glu	Gly	Phe	Asn	11e 400
Ser	Asp	Lys	Asp	Met 405	Glu	Lys	Glu	Tyr	Arg 410	Gly	Gln	Asn	Lys	Ala 415	Ile
Asn	Lys	Gln	Ala 420	Tyr	Glu	Glu	Ile	Ser 425	Lys	Glu	His	Leu	Ala 430	Val	Tyr
Lys	Ile	Gln 435	Met	Cys	Lys	Ser	Val 440	Lys	Ala	Pro	Gly	Ile 445	Cys	Ile	Asp
Val	Asp 450	Asn	Glu	Asp	Leu	Phe 455	Phe	Ile	Ala	Asp	Lys 460	Asn	Ser	Phe	Ser
Asp 465	Asp	Leu	Ser	Lys	Asn 470	Glu	Arg	Ile	Glu	Tyr 475	Asn	Thr	Gln	Ser	Asn 480
Tyr	Ile	Glu	Asn	Asp 485	Phe	Pro	Ile	Asn	Glu 490	Leu	Ile	Leu	Asp	Thr 495	Asp
Leu	Ile	Ser	Lys 500	Ile	Glu	Leu	Pro	Ser 505	Glu	Asn	Thr	Glu	Ser 510	Leu	Thr
qaA	Phe	Asn 515	Val	Asp	Val	Pro	Val 520	Tyr	Glu	Lys	Gln	Pro 525	Ala	Ile	Lys
Lys	Ile 530	Phe	Thr	Asp		Asn 535	Thr	Ile	Phe	Gln	Tyr 540	Leu	Tyr	Ser	Gln
Thr 545	Phe	Pro	Leu												

PCT/GB97/02273 WO 98/07864

- 105 -

## (2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2616 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double
    (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION:1..2616

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly		48
				TAC Tyr													9,6
				AAG Lys													144
				AAC Asn												,	192
				CCA Pro													240
				GAT Asp 85												:	288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val		336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys		384
				AAC Asn												•	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	•	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	!	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	!	576

ACG Thr	TTC Phe	GGT Gly 195	Phe	GAC Glu	G GAG	AGC Ser	CTG Leu 200	Glu	GTT Val	GAT Asp	ACC	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	Gly	Lys	TTC Phe	GCA Ala	ACT Thr 215	Asp	CCA Pro	GCG Ala	'GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	Ile	CAC	GCC	GGT	CAT His 230	Arg	CTG	TAT	GGC	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	, 864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
												TTT Phe				960
												GTA Val				1008
												TAC Tyr				1056
												ACA Thr 365				1104
												AAG Lys				, 1152
												TTA Leu				1200
												TTT Phe				1248
AAA Lys												CTA Leu				1296
						Thr						GGA Gly 445				1344
GCA Ala												GAC Asp				1392

	Pro			A GAT Asp		Phe					Asn				GAA Glu 480		1440
				ACT Thr 485	Asn												1488
				CAA Gln					Phe								1536
GAA Glu	AAT Asn	ATT Ile 515	Ser	ATA	GAA Glu	AAT Asn	CTT Leu 520	TCA Ser	AGT	GAC Asp	ATT Ile	ATA Ile 525	GGC Gly	CAA Gln	TTA Leu	•	1584
				AAT Asn													1632
				ACT Thr													1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu		1728
				CGT Arg												•	1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu		1824
				GAT Asp													1872
				GAT Asp												,	1920
TTA Leu	AAT Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	GAT Asp 650	GAT Asp	TTT Phe	GTA Val	GGT Gly	GCT Ala 655	TTA Leu		1968
ATA Ile	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT	GCA Ala	٠	2016
ATA Ile	CCT Pro	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys		2064
Val	CTA Leu 690	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	AAT Asn	GAA Glu		2112
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720		2160
GTT Val	AAT Asn	ACA Thr	Gln	ATT Ile 725	GAT Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu		2208

GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp	2304
			AAA Lys													2352
			TTG Leu													2400
			GGT Gly													2448
			TTA Leu 820												GGT Gly	2496
			AGA Arg			Asp										2544
			CAG Gln													2592
			GAA Glu				TAA •									2616

### (2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 872 amino acids
  - (B) TYPE: amino acid (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Il Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 185 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 280 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 375 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 455

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490' Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 535 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 585 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 650 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 695 Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
705 710 715 720 Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys

WO 98/07864 PCT/GB97/02273

- 111 -

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855 860

Thr Phe Thr Glu Tyr Ile Lys 865 870

- (2) INFORMATION FOR SEQ ID NO: 27:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2574 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
  - (xi) SEQUENCE DESCRIPTION: SEO ID NO: 27:

ATGCCGGTTA CCATCAACÁA CTTCAACTAC AACGACCCGA TCGACAACAA CAACATCATC 60 ATGATGGAAC CGCCGTTCGC ACGTGGTACC GGTCGTTACT ACAAGGCTTT CAAGATCACC 120 GACCGTATCT GGATCATCCC GGAACGTTAC ACCTTCGGTT ACAAACCTGA GGACTTCAAC 180 AAGAGTAGCG GGATTTTCAA TCGTGACGTC TGCGAGTACT ATGATCCAGA TTATCTGAAT 240 ACCAACGATA AGAAGAACAT ATTCCTTCAG ACTATGATCA AGTTATTTAA TAGAATCAAA 300 TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA 360 GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA 420 :TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAAA TTTAATAATA 480 TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAAATCAT 540 TTTGCATCAA GGGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTTGCCC AGAATATGTA 600 AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACGTGGATAT 660 TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTTATAC ATGTTTTACA TGGATTATAT 720 GGCATTAAAG TAGATGATTT ACCAATTGTA CCAAATGAAA AAAAATTTTT TATGCAATCT 780 ACAGATGCTA TACAGGCAGA AGAACTATAT ACATTTGGAG GACAAGATCC CAGCATCATA 840 ACTCCTTCTA CGGATAAAAG TATCTATGAT AAAGTTTTGC AAAATTTTAG AGGGATAGTT 900 GATAGACTTA ACAAGGTTTT AGTTTGCATA TCAGATCCTA ACATTAATAT TAATATATAT 960 AAAAATAAAT TTAAAGATAA ATATAAATTC GTTGAAGATT CTGAGGGAAA ATATAGTATA 1020 GATGTAGAAA GTTTTGATAA ATTATAAAA AGCTTAATGT TTGGTTTTAC AGAAACTAAT 1080 ATAGCAGAAA ATTATAAAAT AAAAACTAGA GCTTCTTATT TTAGTGATTC CTTACCACCA 1140 GTAAAAATAA AAAATTTATT AGATAATGAA ATCTATACTA TAGAGGAAGG GTTTAATATA 1200

	TCTGATAAAG	ATATGGAAAA	AGAATATAGA	GGTCAGAATA	AAGCTATAAA	TAAACAAGCT	1260
	TATGAAGAAA	TTAGCAAGGA	GCATTTGGCT	GTATATAAGA	TACAAATGTG	TAAAAGTGTT	1320
	AAAGCTCCAG	GAATATGTAT	TGATGTTGAT	AATGAAGATT	TGTTCTTTAT	AGCTGATAAA	1380
	AATAGTTTTT	CAGATGATTT	ATCTAAAAAC	GAAAGAATAG	AATATAATAC	ACAGAGTAAT	1440
	TATATAGAAA	ATGACTTCCC	TATAAATGAA	TTAATTTTAG	ATACTGATTT	AATAAGTAAA	1500
	ATAGAATTAC	CAAGTGAAAA	TACAGAATCA	CTTACTGATT	TTAATGTAGA	TGTTCCAGTA	1560
	TATGAAAAAC	AACCCGCTAT	AAAAAAAATT	TTTACAGATG	AAAATACCAT	CTTTCAATAT	1620
	TTATACTCTC	AGACATTTCC	TCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
	GATGCATTAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTT	CTATGGATTA	TATTAAAACT	1740
	GCTAATAAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
	TTTGTAATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
	GTTCCTTATA	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
	AATGCTTTTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
	CCTGTAGTTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	АТАААААТАА	AATTATTAAA	2040
	ACAATAGATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	,2100
	GTAGCGCAAT	GGCTCTCAAC	-AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
	AAGGCTTTAA	ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	АТАТААТАТА	2220
	TATTCTGAAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280
	AATGAGGGTA	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
	TCATATTTAA	TGAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
•	ACTOTOAAAA	AAAATTTGTT	ATATATAAA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
	GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
	TCAATATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

#### (2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2574 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double

  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

ATGCCAGTTA CAATAAATAA TTTTAATTAT AATGATCCTA TTGATAATAA TAATATTATT 60 ATGATGGAGC CTCCATTTGC GAGAGGTACG GGGAGATATT ATAAAGCTTT TAAAATCACA 120 CATCGTATTT GGATAATACC GGAAAGATAT ACTTTTGGAT ATAAACCTGA GGATTTTAAT 180 AAAAGTTCCG GTATTTTTAA TAGAGATGTT TGTGAATATT ATGATCCAGA TTACTTAAAT 240 WO 98/07864 PCT/GB97/02273

- 113 -

ACTAATGAT	A AAAAGAATA	T ATTTTTACA	A ACAATGATC	A AGTTATTA	TAGAATCAAA	. 300
TCAAAACCA:	r tgggtgaaa	a gttattaga	G ATGATTATA	A ATGGTATACO	TTATCTTGGA	360
GATAGACCTO	3 TTCCACTCG	A AGAGTTTAA	C ACAAACATTO	G CTAGTGTAAC	TGTTAATAAA	420
TTAATCAGTA	ATCCAGGAGA	A AGTGGAGCG	A AAAAAAGGTA	TTTTCGCAAA	TTTAATAATA	480
TTTGGACCT	GGCCAGTTT	r aaatgaaaa:	GAGACTATAC	S ATATAGGTAI	ACAAAATCAT	540
TTTGCATCA	GGGAAGGCTT	CGGGGGTATA	A ATGCAAATGA	AGTTTTGCCC	AGAATATGTA	600
AGCGTATTTA	ATAATGTTCA	A AGAAAACAA	A GGCGCAAGTA	\ TATTTAATAG	ACGTGGATAT	660
TTTTCAGATO	CAGCCTTGAT	T ATTAATGCAT	GAACTCATCC	ACGTCCTCCA	CGGTCTCTAC	720
GGTATCAAAG	TAGACGACCT	CCCGATCGT	CCGAACGAAA	AAAAATTCTT	CATGCAGAGC	780
ACCGACGCAA	TCCAGGCAGA	AGAACTCTAC	ACCTTCGGTG	GTCAGGACCC	GAGCATCATC	840
ACCCCGAGCA	CCGACAAAAG	CATCTACGAC	AAAGTCCTCC	AGAACTTCCG	TGGTATCGTC	900
GACCGTCTCA	ACAAAGTCCT	CGTCTGCATC	AGCGACCCGA	ACATCAACAT	CAACATCTAC	960
AAAAACAAAT	TCAAAGACAA	ATACAAATTC	GTCGAAGACA	GCGAAGGTAA	ATACAGCATC	1020
GACGTCGAGA	GCTTCGACĀA	ACTCTACAAA	AGCCTCATGT	TCGGTTTCAC	CGAAACCAAC	1080
ATCGCAGAAA	ACTACAAAAT	CAAAACCCGT	GCAAGCTACT	TCAGCGACAG	CCTCCCGCCG	, 1140
GTCAAAATCA	AAAACCTCCT	CGACAACGAA	ATCTACACCA	TCGAAGAAGG	TTTCAACATC	1200,
AGCGACAAAG	ACATGGAAAA	AGAATACCGT	GGTCAGAACA	AAGCAATCAA	CAAACAAGCT	1260
TACGAAGAAA	TCAGCAAAGA	ACACCTCGCA	GTCTACAAAA	TCCAGATGTG	CAAAAGCGTC	1320
AAAGCACÇGG	GTATCTGCAT	CGACGTTGAC	AACGAAGACC	TCTTCTTCAT	CGCAGACAAA	1380
AACAGCTTCA	GCGACGACCT	CAGCAAAAAC	GAACGTATCG	AATACAACAC	CCAGAGCAAC	1440,
racatcgaaa	ACGACTTCCC	GATCAACGAA	CTCATCCTCG	ACACCGACCT	CATCAGCAAA	1500
ATCGAACTCC	CGAGCGAAAA	CACCGAAAGC	CTCACCGACT	TCAACGTTGA	CGTCCCGGTC	1560
TACGAAAAAC	AGCCGGCAAT	CAAAAAAATC	TTCACCGACG	AAAACACCAT	CTTCCAGTAC	1620
TCTACAGCC	AGACCTTCCC	GCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
SATGCATTAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTTT	CTATGGATTA	TATTAAAACT	1740
CTAATAAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTGTAATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
TTCCTTATA	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
ATGCTTTTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CTGTAGTTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	ATAAAAATA	AAATTATTAAA	2040
CAATAGATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
TAGCGCAAT	GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AGGCTTTAA	ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	ATATAATATA	2220
ATTCTGAAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280

PCT/GB97/02273

- 114 -

AATGAGGGTA	TTAACCAAGC	TATAGATAAT	TTAATAATT	TTATAAATGG	ATGTTCTGTA	2340
<b>ICATATTTAA</b>	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA	AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG	AAAAAȚCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	CTTTGATCTT	2520
CAATATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

### **CLAIMS**

- 1. A polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and wherein said second domain is adapted (i) to translocate the polypeptide into a cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into a cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of clostridial neurotoxin precursor that can be converted into toxin by proteolytic action.
- 2. A polypeptide according to Claim 1 wherein said first domain comprises a clostridial toxin light chain.
- A polypeptide according to Claim 1 wherein said first domain comprises a fragment or variant of a clostridial toxin light chain.
- 4. A polypeptide according to Claim 2 or 3 wherein the clostridial toxin is a botulinum toxin.
- 5. A polypeptide according to any preceding claim wherein the first domain exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
- 6. A polypeptide according to any preceding claim wherein said second domain comprises a clostridial toxin heavy chain  $H_N$  portion.
- 7. A polypeptide according to any of Claims 1-5 wherein said second domain comprises a fragment or variant of a clostridial toxin heavy chain  $H_N$  portion.
- 8. A polypeptide according to Claim 6 or 7 wherein the clostridial toxin is a

botulinum toxin.

- 9. A polypeptide according to any of Claims 1-8 further comprising a third domain adapted for binding of the polypeptide to a cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that bind to a cell.
- 10. A polypeptide according to Claim 9 wherein said third domain is for binding the polypeptide to an immunoglobulin.
- 11. A polypeptide according to Claim 10 wherein said third domain is a tandem repeat synthetic IgG binding domain derived from domain  $\beta$  of Staphylococcal protein A.
- 12. A polypeptide according to Claim 9 wherein said third domain comprises an amino acid sequence that binds to a cell surface receptor.
- 13. A polypeptide according to Claim 12 wherein said third domain is insulin-lik growth factor-1 (IGF-1).
- 14. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and a portion designated  $H_N$  of a botulinum toxin heavy chain.
- 15. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A.
- 16. A polypeptide according to Claim 15 wherein the botulinum toxin type A light chain variant has at residue 2 a glutamate, at residue 26 a lysine and at residue 27 a tyrosine.

PCT/GB97/02273

- 117 -

WO 98/07864

- 17. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
- 18. A polypeptide according to any of Claims 1-13 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.
- 19. A polypeptide according to Claim 18 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 20. A polypeptide according to Claim 15 or 16 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
- 21. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 22. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 23. A polypeptide according to Claim 17 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
- 24. A polypeptide according to Claim 23 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 25. A polypeptide according to any of Claims 14-24 lacking a portion designated

H<sub>c</sub> of a botulinum toxin heavy chain.

- 26. A polypeptide comprising a botulinum toxin light chain and a fragment of a botulinum toxin heavy chain, said fragment being not capable of binding to cell surface receptors.
- 27. A polypeptide according to Claim 26 lacking an intact portion designated  $H_{\rm c}$  of a botulinum toxin heavy chain.
- 28. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
- 29. A polypeptide according to Claim 28 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
- 30. A polypeptide according to Claim 28 or 29 comprising a variant of a clostridial toxin heavy chain  $H_N$  portion and further comprising a site for cleavag by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain  $H_N$  portion.
- 31. A polypeptide according to Claim 28, 29 or 30 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotid s coding for the cleavage site.
- 32. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-31 with (b) a second polypeptide being a polypeptide or oligopeptide adapted for binding to an affinity matrix so as to enable purification of the fusion protein using said matrix.
- 33. A fusion protein according to Claim 32 wherein said second polypeptide is

WO 98/07864 PCT/GB97/02273

adapted to bind to a chromatography column, such as an affinity matrix of glutathione Sepharose.

- 34. A fusion protein according to Claim 32 or 33 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.
- 35. A composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the botulinum toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*.
- 36. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a positive control in a toxin assay.
- 37. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a vaccine against clostridial toxin.
- 38. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for *in vivo* use.
- 39. A pharmaceutical composition comprising a composition according to Claim 35, a polypeptide according to any of claims 1-31 or a fusion protein according to Claim 32, 33 or 34, in combination with a pharmaceutically acceptable carrier.
- 40. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-34.
- 41. A nucleic acid encoding a polypeptide or a fusion protein according to Claim



WO 98/07864 PCT/GB97/02273

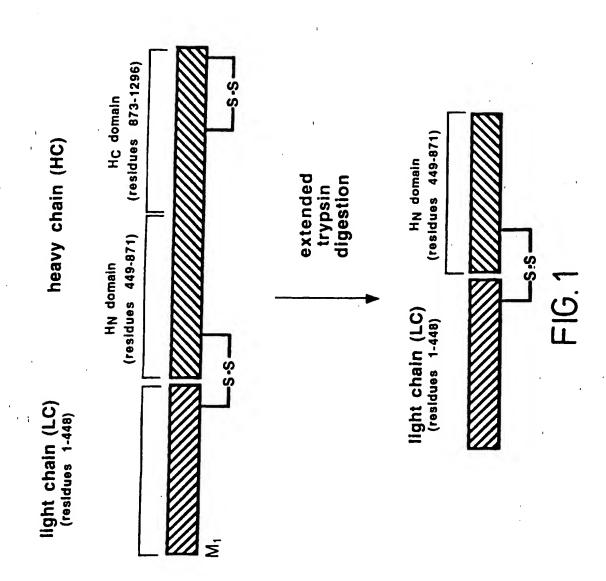
- 120 -

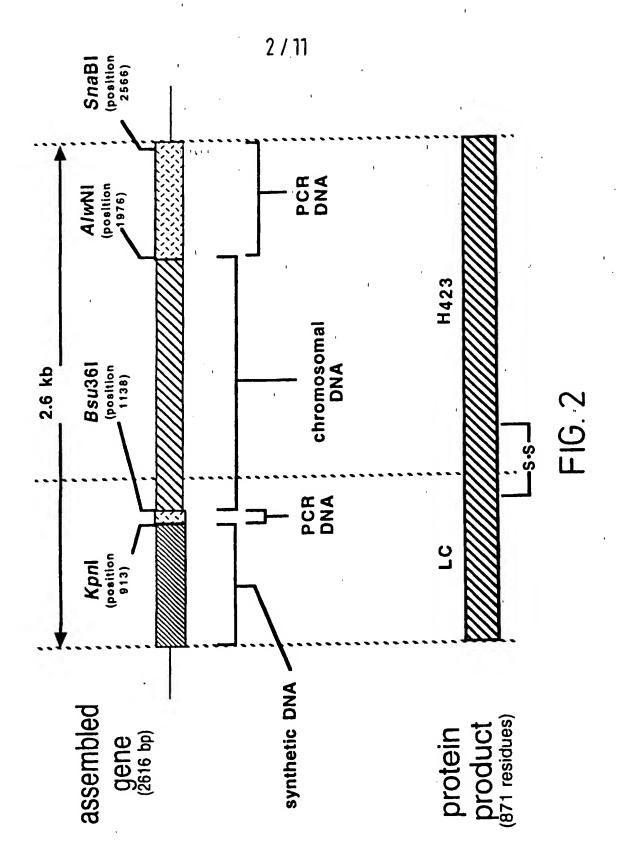
40 and comprising nucleotides encoding residues 1-448 of a botulinum toxin type A light chain.

- 42. A nucleic acid according to Claim 40 or 41 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain  $H_N$  domain.
- 43. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 and comprising nucleotides encoding residues 1-470 of a botulinum toxin typ B light chain.
- 44. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 or 43 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain  $H_N$  domain.
- 45. A nucleic acid according to any of Claims 40-44 comprising nucleotides encoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.
- 46. A nucleotide according to Claim 45 obtainable by modification of a nucleotide encoding a polypeptide or fusion protein according to any of claims 1-34 so as to introduce said cleavage site.
- 47. A DNA according to any of claims 40-46.
- 48. A DNA selected from SEQ ID No:s 1, 8, 10, 12, 14, 16, 18, 23 and 24.
- 49. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid according to any of Claims 40-48 and recovering the polypeptide.
- 50. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid encoding a fusion protein

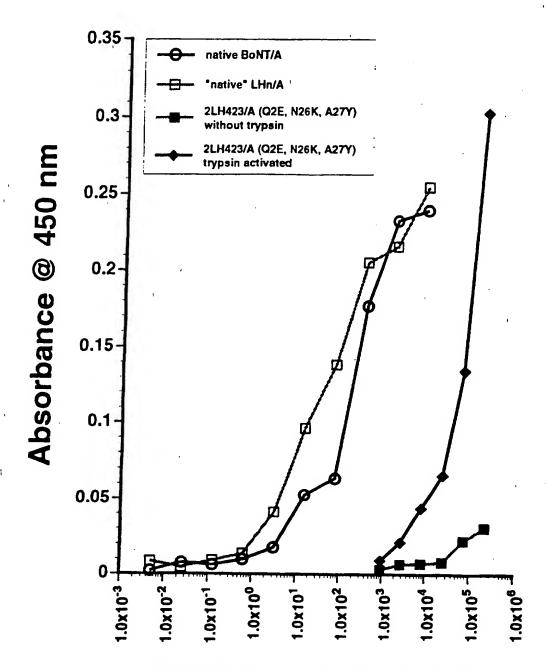
according to Claim 32, 33 or 34, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.

- 51. A method of manufacture according to Claims 49 or 50 in which the nucleic acid is DNA.
- 52. A cell expressing a polypeptide or fusion protein according to any of Claims 1-34.





.....

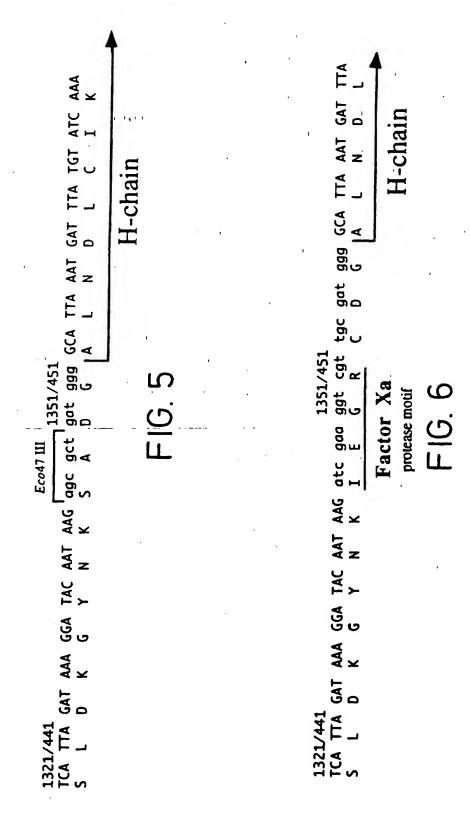


Protein concentration (ng/ml)

FIG. 3

. 2)		<del>5</del>	4/11 ©		
QFNYKDPVNGVDIAYIKIPNAGQMQPV (SeqID. 2)	ľ	Q P V (Seq I.D. 4)	NGVD-IAYIKIPKYGQMQPV (Seq1.D.6)		
>		>	>	>	>
۵		۵	۵	٩	١
0		0	0	0	· 0
Σ		Σ	Σ,	Σ	Σ
0		0	0	0	0
5		_ ق	ی	<u> </u>	ی
4	1	<b>&gt;</b>	<b>&gt;</b>	Σ Ο Δ Ζ	∢
Z	•	IKIPKYGQM	<u>×</u>	Z	QFNYKDPVNGVDIAYIKIPNAGQMQ
۵		٩	<u>م</u>	QFNYKDPVNGVDIAYIKIP	, •
H		H	Н	H	H
<b>×</b> .		×	¥	<b>×</b>	<b>×</b>
Η .			H	H	H
<b>&gt;</b>		>	<i>-</i>	_	-
⋖ ,	•	DIAY	1		
H	۵.		-1		
	EL	>	<u> </u>	>	>
7	ш	9	5	ی	ی
7	ш	z	z	z	z
>	STRLQKLLEF	>	>	>	>
_		> 4	QFNYKDPV	٥	<u> </u>
۵	¥	0	0	۵	0
¥	0	¥	<b>×</b> .	¥	×
<b>&gt;</b>		<b>&gt;</b>	>	<b>&gt;</b>	<b>&gt;</b>
z	œ	Z	Z	Z	Z
LL	-	Q F N Y	ட	ш.	u_
	S				
¥	<b>-</b>	×	×	×	×
Z	X	Z	Z	22	Z
>	I	>	>	>	>
MQF V N K	G S P G I H M T	G T M E F V N K	G S M E F V N K 123	M	MPFVNK
<u> </u>	6	<u></u>	<del></del>	<del></del>	
<b>∑</b> ⊣	10	T M 24	G S M 123	<b>2</b> 17	<b>~</b> ***
	(5 -4		2 7		
			•		
LH423/A .	H <sub>423/A</sub>	(Q <sub>2</sub> E, N <sub>26</sub> K, A <sub>27</sub> Y)	<sub>2</sub> LH <sub>423</sub> /A (Q <sub>2</sub> E, N <sub>26</sub> K, A <sub>27</sub> Y)	Native BoNT/A, C. botulinum 2169 Thompson et al.1990	Native BoNT/A, C. <i>botulinum</i> 62A Binz <i>et al.</i> 1990
آ د	236	Ö	, , , , , , , , , , , , , , , , , , ,	Nativ C. bc Thor	Nativ C. bx Binz

= REGIONS OF NON-IDENTITY WITH THE NATIVE SEQUENCES.



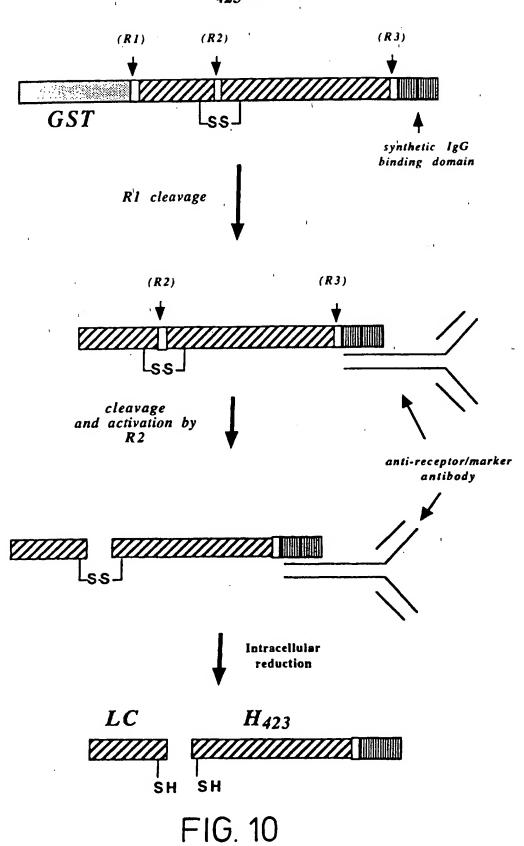
AGG R ACA T TAT. GAC D CAG 0.: ATG M TCT AGG GCG CCT A P 66A 6 GAG E 151 CTG L AGG R AAG K GTG V AGG R ر 8 11C F CTA AGG TAT Y AGT S CAG Q AGC S GAA E GTG V GAA E ACA T AGC S TAT Y 999 TCT S CTG L CGG GCT A TCA S T L ACA T GAG E Ę, AAG K T A ٥ GCT A 75 75 70 9 8 8 AGA R 999 AAG K 5 5 7 GAG E CCT & o AAC N 750 AAG K F 7 AAT N GAT D CTC 2587/863 TAC GTA GAT A Y V D N 2647/883 CCG GAG ACG ( P E T I 2707/903 GGC TTT TAT G G F Y I C F Y I

TCT AGG CCT | S R P -ACA CAT AAT T H N ATT AAG I K GAT D ATT TAT Y GAT D 2617/873 TTT ACT GAA T F T E Y 2677/893 TAT CAA TCT ( TCT ACA S T ပ္ပိုင္ပ TCA S TTA L E 4 TT P ATA I AGA R **₹**~ AGA R S & T L E SA AAA K AAT N 2587/863 TAC GTA GAT / Y V D I 2647/883 TCT AAA GTT / S K V I 2707/903 ATT AAG GAT I

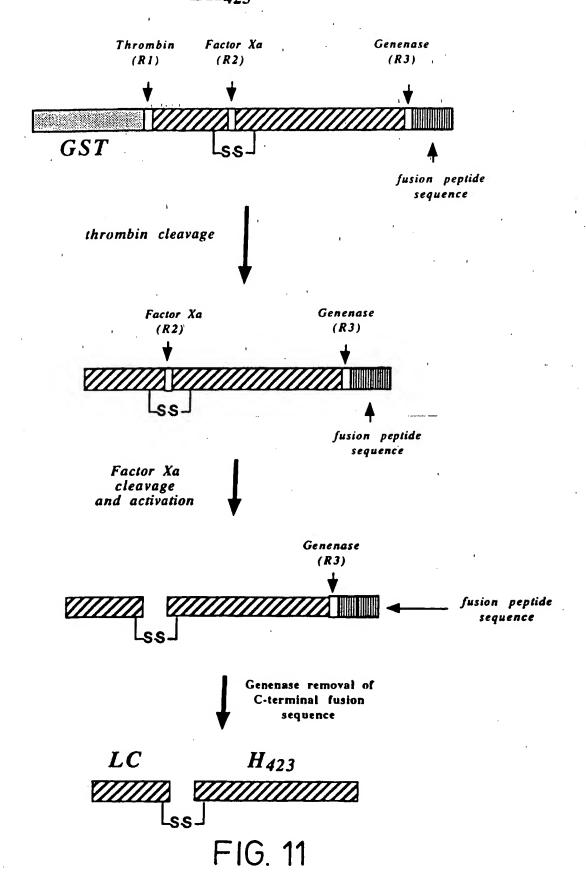
F1G. 8

AA N	AAA K	CGA R	GAA E	CAA 0	90CC	AAA
						GCT A
9	TTC F	GAA	TTA L	AAA	CGA,	GAA
TCA S	AAA K	GAA E	CTT 1	AAC	CAA Q	GCA A
		AAC				,
		TTA L				
TAT Y	GTA V	AAC	AGC S	AAC N	AAC	N N
GAA Ē	900 A	CCT P	CAA Q	GAC	TTA L	CAA AGC GCT Q S A 3037/1013 GAC TAG D *
7/873 ACT T	GAA E	TTA L /933	AGC 5 5	GTA V 7973	AAC N	AGC S S 17/101 TAG
2617 TTT F 2677	6AT D 2737	CAT H 2797	CCA P 2857	AAA K	CCT P 2977	CAA 0 3037 6AC
ACA T	CAC	TTA	SAC	ررو م	TTA 'L	AGC S GTA V
TCT S	Q Q	ATC I	GAT D	900 A	CAT H	CCA AAA K
TTA	900 A	GAG E	X A A A	CAG	TTA L	SAC D CCG
						GAT D GCG A
AGA R	¥ ±	T F	AGT S	GAT D	GAG E	AAA CAG
CAA	₽ A	6CG A	<b>\$</b>	N N	TAT Y	L L GCT A
A -	5	A -	) I	TA	۲	AT (
GAT D	50	O CAA	77C F	AAG K	6CG A	CAA AAT
/863 GTA V /883	CCG	CAA 0 7923	6CC A 7943	AAA K	AAC N	ATC I 7/100 CTA
2587 TAC Y 2647	TCC S 2707	GAA E 2767	AAC N 2827	GCT A 2887	CAA 2947	TTC ATC CAA A 3007/1003 AAG CTA AAT C

# 8/11 LH<sub>423</sub>/A



# $LH_{423}/A^{9/11}$



10/11

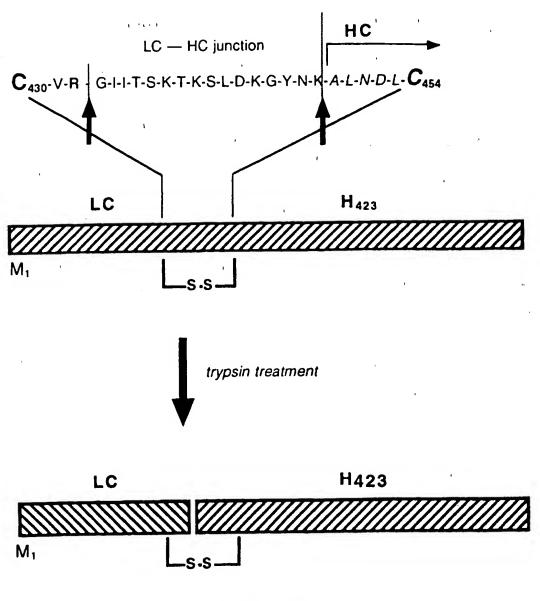
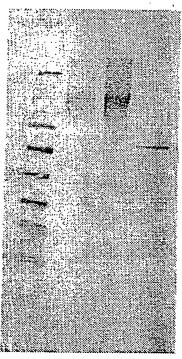


FIG. 12

11 / 11

Panel A. 1 2 3 4



Panel B. 1 2 3 4

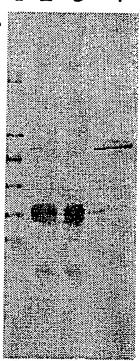


FIG. 13

### INTERNATIONAL SEARCH REPORT

Inten unal Application No PCT/GB 97/02273

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/31 C12N C07K14/33 A61K38/16 C12N1/21 C12P21/02 A61K39/08 According to International Patent Classification (IPC) or to both national classification and IPC B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) C12N C12P A61K IPC 6 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practical, search terms used) C. DOCUMENTS CONSIDERED TO BE RELEVANT Relevant to claim No. Citation of document, with indication, where appropriate, of the relevant passages Category \* 1-52 X WO 96 12802 A (OPHIDIAN PHARM INC :WILLIAMS JAMES A (US); PADHYE NISHA V (US); KI) 2 May 1996 see the whole document 1-52 KURAZONO H ET AL: "Minimal essential . X \*domains\* specifying toxicity of the \*light\* \*chains\* of tetanus toxin and botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 (21) P14721-9, UNITED STATES, XP002047910 see table II Patent family mombers are listed in annex: Further documents are listed in the continuation of box C. X Special categories of cited documents "T" later document published after the international filing date or priority date and not in conflict with the application but "A" document defining the general state of the art which is not cited to understand the principle or theory underlying the considered to be of particular relevance "E" earlier document but published on or after the international \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to filing date involve an inventive step when the document is taken alone "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means document published prior to the international filing date but later than the priority date claimed "&" document member of the same patent family Date of mailing of the international search report Date of the actual completion of the international search 3 0. or 98 9 December 1997 Authorized officer Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentinan 2 NL - 2280 HV Rqswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo ni, Hillenbrand, G Fax: (+31-70) 340-3016

### INTERNATIONAL SEARCH REPORT

Inten .onal Application No PCT/GB 97/02273

	•	PCT/GB 9	//022/3
C.(Continue	ation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages		Relevant to claim No.
x	LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938 see the whole document		1
A	BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348 see the whole document		1,26,35
			1
			1
	·		
	1		

## INTERNATIONAL SEARCH REPORT

tntei. .onal Application No PCT/GB 97/02273

, Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9612802 A	02-05-96	US 560182		11-02-97
		US 5196193	3 A	23-03-93
		AU 396839!	5 A	15-05-96
		EP 0796320	5 A	24-09-97
•		FI 971732	2 A	23-06-97
		NO 971868	3 A	24-06-97
•		PL 320214	I A	15-09 <b>-</b> 97
		ZA 9508990	) A	15-05-96
		AU 6653894	A	04-07-94
		CA 2150935	5 A	23-06-94
		EP 0671902	2 A	20-09-95
		WO 9413264	A	23-06-94
		US 5466672	2 A	14-11-95
		US 5599539		04-02-97
ı		AU 638786		08-07-93
		AU 6895191	L A	31-05-91
		EP 0498854	A	19-08-92
		WO 9106306		16-05-91
		US 5443976		22-08-95
		US 5340923		23-08-94

Form PCT/ISA/210 (patent family annex) (July 1992)